The Impact of Breathiness on Speech Intelligibility in Pathological Voice

A thesis submitted in partial fulfilment of the requirements for the degree of

Master of Audiology

in the Department of Communication Disorders

at the University of Canterbury

By

Louise Thompson

University of Canterbury

2011

Table of Contents

Acknowledgements

I would like to acknowledge and thank my supervisor Dr. Emily Lin for her support, advice, patience and time spent helping me with the planning, data collection, statistics, and writing of this thesis. Emily's guidance has been indispensable. I would like to thank my participants for their involvement in this study. I also wish to thank my classmates and family who have provided support and encouragement.

Abstract

Aim

The aim of this study was to determine how deterioration of voice quality, such as breathiness, may impact on the intelligibility of speech.

Method

Acoustic analysis was conducted on sustained vowel phonation (/i/ and /a/) and sentences produced by voice disordered speakers. Measures included: frequency and amplitude of the first two formants (F1, F2), singing power ratio (SPR), the amplitude difference between the first two harmonics (H1-H2), voice onset time (VOT), and energy ratio between consonant and vowel (CV energy ratio). A series of two-way (glottal closure by vowel) mixed design between and within-subjects Analysis of Variances conducted on these acoustic measures showed a significant glottal closure (complete and incomplete) or glottal closure by vowel interaction effect on the F2 frequency, H1-H2 amplitude difference, and singing power ratio. Based on findings in literature that reported a dominant first harmonic as a useful predictor of breathiness, the measure of H1-H2 amplitude difference was selected as a factor for investigation of the impact of voice quality on the perception of vowel intelligibility and clarity. Fixed-length vowel segments at five levels of H1-H2 amplitude difference were presented to 10 male and 10 female inexperienced listeners between the ages of 19 and 34 years.

Results

It was expected that the tokens with a dominant first harmonic, indicative of a more breathy voice, would be associated with a lower rate of correct vowel identification and of being perceived as "clearer". Although no linear relationship between breathiness and intelligibility was revealed, results indicated the presence of thresholds of intelligibility for

particular vowels whereby once a level of breathiness was reached intelligibility would decline.

Conclusion

The finding of a change of the perceptual ratings as a function of the H1-H2 amplitude difference, identified in previous studies as a measure of breathiness, revealed thresholds of intelligibility for particular vowels below which breathiness would be tolerated with little impact on intelligibility but beyond which intelligibility ratings suffered markedly.

List of Tables

List of Figures

1 Introduction

1.1 Thesis Overview

Speech intelligibility is key to oral communication and thus of main concern in the field of communication disorders and sciences. Voice disorders, for example, can disrupt voice quality in ways that undermine speech intelligibility. To date, there are many perceptual, as well as acoustic, studies of voice quality and speech intelligibility in the literature. However, there is a paucity of direct investigations on the relationship between voice quality and speech intelligibility. To provide the empirical data that would help identify which voice-related changes may affect speech intelligibility, the present study examines what effect a decline in voice quality, such as increased breathiness, would have on the intelligibility of speech. This study includes two parts of investigation, one using an acoustic and the other a perceptual approach. The first part of the study compares two groups of voice patients on a variety of acoustic measures extracted from segments embedded in sentences and sustained vowels produced by these patients. Voices produced by voice patients found to show a complete glottal closure during phonation under videolaryngostroboscopic examinations were compared with voices obtained from voice patients identified as exhibiting an incomplete glottal closure. In the second part of the study, a measure found in the literature to distinguish between breathy and non-breathy voices was used to select voice samples for presentation to a group of normal hearing listeners to determine how acoustic differences related to breathiness may affect the perception of vowel intelligibility and clarity. The purpose of this study is to identify which acoustic measures are sensitive in detecting a difference between pathological voice produced by voice patients with and without complete glottal closure and explore how changes in the acoustic measure of voice quality such as breathiness may impact on the perception of vowel clarity.

1.2 Literature Review

This section provides the background information about the definition of voice quality and breathiness as well as a literature review covering three main areas related to the present investigation. Firstly, the concept of breathiness is examined in relation to pathological as well as normal voices, including the use of breathiness in languages for linguistic contrast and its role in differentiating between male and female voice. Secondly, findings related to the effect of vocal control on speech intelligibility are extrapolated from previous studies of "clear speech" and the speech of the hearing impaired. Thirdly, a critical review of the literature is included to clarify the current state of the understanding about the acoustic measures that have featured in studies concerning breathiness as well as those employed in the present investigation of voice quality and speech intelligibility.

1.2.1 Voice Quality and Breathy Voice

Voice quality refers to the perceived auditory characteristics that mark an individual's speech (Gerratt & Kreiman, 2004). It is what remains when other dimensions of voice such as pitch, loudness, and phonetic category have been excluded (Titze, 2000). It includes characteristics such as roughness, breathiness, creakiness, and nasality (Titze, 2000). Voice quality is what most concerns people with voice disorders (Kreiman, Gerratt, Kempster, Erman, & Berke, 1993 236). Voice patients seek treatment because they do not sound normal and may thus often judge the success of treatment on whether they sound better (Kreiman et al., 1993). Voice specifically refers to the sound produced at the glottis by vocal fold vibration (Titze, 1994). Vocal folds are muscularized mucosal folds that project from the lateral walls of laryngeal cartilage at the narrowest portion of the airway (Sapienza & Ruddy, 2009). Vocal folds vibrate, colliding at rapid rates when air from the lungs rushes past through the glottis (the air space between them) creating voice (Titze, 1994). In the production of breathy voice, the vocal folds vibrate but, unlike normal phonation, are never closed along their full length (Reetz & Jongman, 2009). The lack of interruption to airflow gives the voice a breathy or husky quality (Reetz & Jongman, 2009). The additional noise associated with breathy voice is attributed to the air turbulence, called aspiration, when it coincides with voicing (Titze, 1994).

1.2.1.1 Voice Quality in Pathological Voice

Pathological voice refers to the voice characterized by signs associated with voice disorders resulting from mass lesions, organic (non-mass lesions), neurological problems, or functional problems (Sapienza & Ruddy, 2009). Mass lesions of the vocal folds that result in voice disorders include Reinke's oedema, vocal fold nodules and cysts, vocal fold thickening, and granuloma, amongst others (Sapienza & Ruddy, 2009). Examples of non-mass lesions that cause voice symptoms are sulcus vocalis, vocal fold bowing (atrophy), asthma, and reflux laryngitis (Sapienza & Ruddy, 2009). Neurological damage can result in dysarthrias, which are speech disorders caused by weakness, delay, and incoordination of sub-glottal, laryngeal, and supralaryngeal systems (Sapienza & Ruddy, 2009). Head and neck tension is a functional cause that can result in a type of strained, high pitched dysphonia (Sapienza & Ruddy, 2009). Across different types of voice problems, common symptoms include breathiness, hoarseness, roughness, strain, asthenia (weakness), diplophonia (a rough quality related to the simultaneous production of two frequencies), voice breaks, pitch breaks, and other anomalies in pitch (Sapienza & Ruddy, 2009). In the voice literature, "breathy" is amongst the most common labels used in describing pathological voice (Kreiman, Gerratt, & Berke, 1994).

1.2.1.2 Breathiness in Normal Voice

Although breathiness is symptomatic of a wide range of pathological conditions, it is also found in non-pathological voice. In fact, breathy sounds feature routinely in some languages. Moreover, breathiness has been found in some studies to characterise normal speaking female voices.

1.2.1.2.1 Breathiness as a Feature in Languages

Ladefoged (1983) noted that pathological voice qualities of English are employed for phonological contrast in other languages. The normal presence of breathiness in a language enables a phonation type to be better studied because its use by all speakers in a language is consistent unlike in a clinical population where its use is more likely to change from instant to instant (Ladefoged, 1983). Breathiness or murmur is used non-contrastively in the English intervocalic [h] in "behind" and "ahead" and contrastively in Indo-Aryan languages such as Nepali, Hindi, and Gujerati and African languages such as West African language Igbo and !Xóõ, a language of South African Bushmen (Ladefoged, 1983). !Xóõ distinguishes normally voiced and murmured vowels (Ladefoged, 1983). The contrast between breathy and clear vowels is achieved at the glottis seemingly by altering the larynx configuration (Bickley, 1982).

In a study of the speech recorded from ten speakers of !Xóõ who were instructed to say six words, Ladefoged (1983) found that the amplitude difference between the largest harmonic of the first formant (F1) and fundamental frequency (F0), also referred to as the first harmonic (H1), was most reliable and statistically significant in differentiating between vowel types. Specifically, the difference in intensity between F0 and the largest harmonic of the first formant was found to be greater for the regularly voiced vowels than the murmured vowels (Ladefoged, 1983). He found jitter and spectral tilt to be unreliable predictors of vowel type (Ladefoged, 1983). A comparison of the spectra of a normally voiced and a murmured vowel showed that the murmured vowel had greater irregular energy in the higher frequencies and less of a falling spectrum (Ladefoged, 1983). The degree of regularity of vocal fold excitation proved difficult to quantify in the vowels because they were short and not produced with a steady F0 (Ladefoged, 1983). Spectral tilt was considered difficult to measure because it was influenced by two opposing factors, one with (1) the less sharp glottal pulses that occur in breathy vowels resulting in a smaller amount of high frequency energy and the other with (2) the greater airflow rates from the turbulence created by an incomplete glottal closure causing greater noise excitation in the higher frequencies (Ladefoged, 1983).

A study by Bickley (1982) confirmed the finding of Ladefoged (1983) in !Xóõ and of Fischer-Jorgensen (1967) in Gujarati that a dominant H1 was consistent with the breathy versions of vowels used for contrast in these two languages. A further study by Huffman (1987) investigated the relationship between laryngeal gestures and voice type using measures of glottal flow and acoustic parameters, including spectral tilt and the amplitude difference between the first two harmonics, to differentiate between vowel produced in Hmong, a Southeast Asian language that uses breathy, normal, and creaky versions of the same vowels for linguistic distinction. Combined oral and nasal airflow was recorded for three Hmong speakers (Huffman, 1987). In addition to spectral measures, inverse filtering was used to recover the glottal flow wave form in the time domain (Huffman, 1987). Both the amplitude difference between H1 and H2 and the closed phase duration ratio (measured from the glottal spectra) are shown to be significant indicators of the differences between phonation types with confirmation from the glottal flow measures (Huffman, 1987). These findings agree with the finding of Bickely (1982) who compared the glottal spectra of vowels used for contrasts in Gujurati and !Xóõ.

1.2.1.2.2 Breathiness in Female Voice

In addition to the use for phonemic discrimination in some languages, breathiness may also be used for gender differentiation. Gender differences in voice exist that can be measured in terms of physiology, acoustics, and perception (Wu & Childers, 1991). The male voice departs from the female voice at puberty due to changes to laryngeal cartilage and vocal folds caused by increased levels of testosterone (Sapienza & Ruddy, 2009). In comparison with female adults, male adults have larger thyroid laminae, a more acute thyroid angle (approximately 90 degrees as opposed to the 120 degrees in women) that results in the larger Adam's apple, thicker vocal folds, and a larger glottal space (Sapienza & Ruddy, 2009). Male laryngeal anatomy grows disproportionately large in relation to the rest of their anatomy (Titze, 2000). It is the bigger mass of the vocal folds that gives men deeper voices (Sapienza & Ruddy, 2009). Adult males have an average speaking fundamental frequency of around 125 Hz compared to an average of 200 Hz for adult females (Titze, 2000). Vocal tract length is also longer in men than in women. The influence of the vocal tract length on voice has been likened to that of a resonator of a musical instrument, with source frequencies generated from the vocal folds and filtered by the vocal tract (Titze, 2000). There is an inverse relationship between vocal tract length and formant frequencies (Titze, 2000). On average, female formants are said to be scaled up in frequency from male formants by about 20% (Wu & Childers, 1991). The gender difference in voice projection power, fundamental frequency, and vocal tract resonance resulting from the physical differences between males and females may also contribute to the difference in voice quality.

As part of a study of variations of voice quality among male and female voices employing acoustic analysis, synthesis, and perceptual measures, Klatt and Klatt (1990) presented natural voice samples produced by 6 male and 10 female normal speaking talkers to a panel of 8 listeners for judgement on a seven-point scale of breathiness (Klatt & Klatt, 1990). It was found that the female voices were perceived on average to be more breathy than male voices (Klatt & Klatt, 1990). However, the range in the levels of breathiness were found to be larger within genders than across genders, with some males being more breathy than many females (Klatt & Klatt, 1990). Therefore, the authors cautioned about making generalisations based on gender or individual behaviour because individuals are likely to produce a range of breathiness with the different speaking styles they adopt in different circumstances (Klatt & Klatt, 1990).

Evidence about the inconsistent relationship between breathiness and female voice can be found in the literature. For example, a Swedish study by Södersten (1995) of 17 healthy women, 45 to 61 years of age, involved airflow, intraoral pressure, and electroglottography measures, and perceptual evaluation. The recorded reading samples were judged by three speech pathologists experienced in perceptual analysis and voice evaluation. In contrast to the Klatt and Klatt's (1990) finding of breathiness in female voices, Södersten (1995) found that female voices were generally non-breathy although they did exhibit a high occurrence of incomplete glottal closure under fiberstroboscopic examination. This finding agrees with the common clinical observation that presence of posterior glottal gap in females is normal and may not necessarily be associated with a breathy voice. However, as women are often found to have larger posterior cartilaginous spaces than men, this anatomical difference between men and women may possibly contribute to a tendency for women to enlarge their glottal space creating breathy voice quality (Sapienza & Ruddy, 2009). The open phase of the glottal pulse is also longer for women allowing greater airflow into the vocal tract causing breathiness (Sapienza & Ruddy, 2009). It is, therefore, not unlikely that due to the anatomical and physiological differences, women are more likely to produce a breathy voice than men.

In addition to the anatomical factors, there have been questions raised as to whether women may adopt breathiness as a functional behaviour. In an investigation into the breathiness of normal female speech, Henton and Bladon (1985) questioned why women would exhibit breathy voices when breathiness is not used contrastively in English and has the potential to reduce speech intelligibility. The authors described breathy phonation as inefficient because it limits vocal intensity, reduces signal-to-noise ratio due to the addition of aspiration noise, and lowers voice pitch, potentially resulting in monotony (Henton & Bladon, 1985). Their hypothesis, informed by offensive sexual stereotype, is that women assume a breathy voice, despite its potential to reduce speech intelligibility, to sound like they are sexually aroused so that they may be perceived as being more desirable (Henton & Bladon, 1985).

Henton and Bladon (1985) studied voice samples from 32 female and 29 male speakers with either of two dialects of British English. The amplitude difference between the first two harmonics (H1-H2) was measured from open vowels /æ, ʌ, ɒ, ɑ/. Open vowels were selected for the study because only they have first formant frequencies high enough not to increase the amplitudes of lower harmonics (Henton & Bladon, 1985). For female voices, averaged across talkers, the first harmonic was higher than the second by 3.3 to 8.4 dB, whereas for male voices the amplitude difference was as small as 0.16 to 0.98 dB (Henton & Bladon, 1985). Although Henton and Bladon's (1985) study did not include perceptual judgements of breathiness nor intelligibility, the acoustic measure alone were taken as evidence of breathiness based on the close correlation between the H1-H2 amplitude difference and the perceptual judgements of breathiness as found in the Bickley (1982) study of breathy vowels used for linguistic contrast.

In summary, as breathy voice is prevalent in both pathological and normal voice, serving as a marker for diagnostic, phonemic, and sociolinguistic discrimination, understanding of its impact on speech intelligibility will help develop useful strategies or designs in speech synthesis, voice analysis or recognition, signal transmission or processing, and speech or voice training.

1.2.2 Speech Intelligibility

Speech intelligibility refers to how easily speech sounds can be decoded (Kent, 1992). It is the behavioural standard by which oral communication is judged (Kent, 1992). Acoustic and perceptual assessments of speech intelligibility are often found in literature investigating the differences between clear and conversational speech and the speech of people with hearing impairment.

1.2.2.1 Intelligibility of Clear versus Conversational Speech

People often adapt their speech to be better understood when confronted with difficult communication environments (Krause & Braida, 2004). This adapted speech style has been named "clear speech". Studies have found substantial gains in intelligibility achieved through use of clear speech in contrast to conversational speech (Picheney, Durlach, & Braida, 1985). Sentence duration was the most obvious difference between clear and conversational speech, with a clear sentence taking roughly twice as long as a conversational sentence (Picheney et al., 1985). A study examining the acoustic characteristics of clear speech compared to conversational speech found that, in clear speech, the durations of phonemes were increased, most consonants had greater intensity, and vowel formants had higher frequencies (Picheny, Durlach, & Braida, 1986). However, as a large number of acoustic differences between clear and conversational speech can be considered (Ferguson &

Kewley-Port, 2007), it remains unclear which acoustic change contributes most to the intelligibility gains.

Krause & Braida (2002), applying appropriate talker training to speakers, managed to elicit clear speech at a normal rate with the intelligibility benefit of slow paced clear speech. A follow-up study (Krause & Braida, 2004) included an analysis and comparison of the properties of clear speech produced at a normal pace (referred to as clear/normal) and conversational speech produced at a normal pace (referred to as conv/normal). The phonetic-level acoustic measures for two of the talkers included vowel formant frequencies, voice onset time (VOT), and consonant-to-vowel (CV) energy ratios, all of which yielded a significant difference between clear and conversational speech except for CV energy ratio (Krause & Braida, 2004). Whilst the finding that the area of the vowel space, as measured from frequencies of the first two formants of the corner vowels, expanded for clear speech produced at a slow pace (referred to as clear/slow) compared to conv/normal speech confirmed findings from Pichney (1986), there was no appreciable expansion in vowel space of clear/normal speech compared to conv/normal speech (Krause & Braida, 2004). The expansion of vowel space, which will be discussed in further detail in Section 1.2.3.2.2 ("Vowel Space"), may result in better vowel differentiation and thus increased speech intelligibility (Liu, Tsao, & Kuhl, 2005). Therefore, this finding suggests that speech rate plays an important role in speech intelligibility, with a slower rate resulting in improved intelligibility.

As for VOT measures, it is generally considered that as speaking rate increases, voice onset time decreases (Baum & Ryan, 1993), coinciding with reduced intelligibility. In Krause and Braida's (2004) study, one talker showed a decrease in VOT for the conv/norm speech compared to the clear/normal speech as was expected while the other talker showed the opposite trend. Contrasting differences in the acoustic properties of the two talkers suggested that they used different strategies to render their speech clear (Krause & Braida, 2004). Significant differences that could contribute to intelligibility gain in clear speech were found with sentence-level measures of long-term spectra, showing increased energy in the 1000-3000 Hz range of the long term spectra, and temporal envelope modulations, showing increased depth of modulation for low modulation frequencies (Krause & Braida, 2004). It was pointed out, however, that the nonsense sentences used as the speech material might lack acoustic properties found in more complex and meaningful speech material (Krause & Braida, 2004). The authors proposed the use of signal processing transformations in further research to endeavour to manipulate acoustic properties in isolation to enable intelligibility testing that evaluates characteristics singly and in combination (Krause & Braida, 2004).

The acoustic characteristics that could account for a gain in speech intelligibility with clear speech have been studied (Picheny et al., 1986). Clear speech was found to have a longer duration than conversational speech, resulting from increased length of speech sounds and insertion of or lengthening of pauses between words (Picheny et al., 1986). It was also found in conversational speech that vowels were modified or reduced and word-final stop consonants were often not released (Picheny et al., 1986). In contrast, it was found in clear speech that there were less modifications of vowels and stop consonants were released as well as essentially all word-final consonants (Picheny et al., 1986). Based on these findings, it appears that the improved speech intelligibility with clear speech, in comparison with conversation speech, is not only related to speech rate but also associated with changes in articulatory and vocal control facilitated by the slowing down of the speech rate.

1.2.2.2 Intelligibility of Hearing-Impaired Talkers

Evidence of the influence of vocal control on speech intelligibility can also be found in investigations conducted on the speech of hearing-impaired talkers. Prelingually deaf speakers without adequate means of amplification are unable to make the same connections between articulation and sensory consequence that mark the speech of normal hearing speakers (Lane &Perkell, 2005). In particular, voicing contrast is often confused or absent (Lane & Perkell, 2005). Poor speech intelligibility ensues such as was shown in an investigation into the phoneme recognition and speech production errors of congenitally hearing impaired children by Smith (1975). With speech samples of 40 children aged between 8 to 15 years with severe-to-profound hearing loss presented to adults who were inexperienced in listening to deaf speech, the intelligibility rating was found to be low, averaging only 18.7% (Smith, 1975). Similarly, a study by John and Howarth (1965) resulted in an intelligibility rating of only 28.6% from speech samples recorded from 29 children aged between 6 to 12 years with severe-to-profound hearing-impairment presented to 20 adult listeners inexperienced at listening to the hearing-impaired.

A study by Monsen (1978) of 67 adolescents with severe-to-profound hearing impairment using markedly shorter and less complex speech materials revealed much higher intelligibility ratings, with an average of 76 %. Two acoustic characteristics were found to correlate highly with intelligibility ratings. These were the difference in VOT used to indicate the distinction between stop consonants /t/ and /d/ and the change in frequency of the second formant (F2) for the distinction between vowels /i/ and /ɔ/ (Monsen, 1978). The explanation given for the F2 frequency being highly correlated with intelligibility rather than first formant (F1) frequency is that variation in the F2 frequency relies more heavily on tongue forwardness whereas large changes in F1 frequency can be achieved with mouth movement (Monsen, 1978). Mouth movement (e.g., lip rounding) can be seen whereas tongue movement is usually a hidden cue for speech readers (Monsen, 1978). In contrast to articulatory parameters, measures of prosody, sentence duration, and the average and extent of variation of F0, were poorly related to speech intelligibility although the author noted that prosody-related measures might be more difficult to quantify or interpret in part because they did not feature as extensively in literature as articulatory measures (Monsen, 1978). Overall, findings about the speech intelligibility of speakers with hearing impairment suggest that aberration in either vocal tract resonance or vocal control may result in reduced speech intelligibility.

1.2.3 Acoustic Measures

As changes in both voice quality and speech intelligibility can be heard, acoustic parameters reflecting these audible changes would provide a link between production and perception. Although acoustic measures do not provide an exact reflection of phonatory process, voice acoustics and vocal fold physiology do correspond and valuable information can be gleaned about phonation from acoustic analysis (Radish Kumar, Bhat, & Prasad, 2009). Physicians use such information to assess changes following treatment (Radish Kumar et al., 2009). The following section is a review of the acoustic measures that have been used to assess breathiness in dysphonic patients and those selected in this study to monitor changes in voice quality and vowel intelligibility.

1.2.3.1 Acoustic Measures Related to Breathiness in Pathological Voice

A normal voice is described as being associated with semi-periodic vocal fold vibration, having little variations in amplitude and frequency between the vibratory cycles of vocal folds for each tone produced. A pathological voice, on the other hand, has greater than normal variations in the periodicity of vocal fold vibration (Hillenbrand, 1988). A breathy voice results from changes to several voicing factors (Klatt & Klatt, 1990). The factors are described by Klatt and Klatt (1990) as an increase in the relative strength of the first harmonic due to an increase in the open quotient, a reduction in the strength of higher harmonics that occurs with an increase in spectral tilt, and the addition of aspiration noise at mid and high frequencies.

Measures designed to capture periodicity include measures of jitter (measuring variability in frequency), shimmer (measuring variability in amplitude), and harmonic-to-noise ratio (HNR) (Radish Kumar et al., 2009). In a study manipulating synthetic signals, Hillenbrand (1987) found that these measures do not operate independently. For example, a large HNR may be due to increased amplitude perturbation or increased pitch perturbation or a combination of the two (Hillenbrand, 1987). Therefore, these perturbation measures are unlikely to indicate the precise source of aperiodicity in voice signals (Hillenbrand, 1987). Furthermore, jitter, shimmer, and HNR were found to become less reliable predictors of dysphonia with greater aperiodicity associated with marked dysphonia (Heman-Ackah et al., 2003). One of the reasons that perturbation measures may be less useful for voice discrimination with severely dysphonic voice is that these measures require accurate tracking of fundamental frequency, which may be challenging when analysing voice with a great amount of aperiodicity (Heman-Ackah et al., 2003; Hillenbrand, 1987; Radish Kumar et al., 2009). This type of concern has resulted in the exclusion of these three measures from one study evaluating acoustic correlates of breathiness (Hillenbrand, Cleveland, & Erickson, 1994).

Cepstral peak prominence (CPP) is a measure that requires a Fourier transformation of the spectrum so that information in the frequency domain of the spectrum is transformed into the time domain and relabelled quefrencies of the cepstrum (Heman-Ackah et al., 2003). The cepstrum shows how frequently frequencies occur in the spectrum (Heman-Ackah et al., 2003). Heman-Ackah et al. (2003) found CPP to be a measure of aperiodicity that has better specificity, sensitivity, and positive and negative predictive value for the measurement of dysphonia than measures of jitter, shimmer, and HNR. The CPP measure, however, is more related to the overall aperiodicity of the voice and has not been associated with any particular type of voice quality.

Fukazawa, el-Assuooty, and Honjo (1988) considered that frequency analysis methods would be more effective than perturbation analysis for measuring aspiration noise because turbulence has a high frequency range independent of vocal pitch range. The measure of spectral tilt is considered useful to reflect the stronger high frequency energy that breathy signals have relative to normal phonation (Hillenbrand et al., 1994). Fukazawa (1988) tested a spectral tilt measure called the Br Index which measures the high-frequency component relative to the total pre-emphasized voice signal. The Br Index was measured from voice samples of 24 normal speaking talkers and 31 talkers who complained of hoarseness and had recently been diagnosed with one of three voice pathologies (10 with vocal cord polyps, 15 with recurrent nerve palsies, and 6 with glottis type laryngeal cancers) (Fukazawa et al., 1988). The voice samples were judged as to the degree of breathiness and roughness they exhibited by two ENT physicians with several years of experience treating voice patients (Fukazawa et al., 1988). The Br Index was found to be significantly correlated with the perception of breathiness and poorly correlated with roughness (Fukazawa et al., 1988). However, Hillenbrand et al. (1994) used a version the Br Index, which was slightly modified from its original version so that it could be measured automatically, to evaluate the ability of acoustic measures to predict breathiness but found it, along with another measure of spectral tilt, to be a weak predictor of breathiness.

A further measure referred to as the high frequency power ratio, which was similar to the Br Index but without the same equipment restrictions required of the measurement of the Br Index, has been proposed for the measurement of breathy voice quality (Shoji, Regenbogen, Jong Daw, & Blaugrund, 1992). The high frequency power ratio is defined as the ratio of high frequency power to the total power (Shoji et al., 1992). In a study investigating how the measure of high frequency power ratio may be used to predict breathiness, recordings of a sustained vowel /ɑ/ obtained from 24 voice patients, whose initial complaint on presentation was breathiness, were compared with those from a normal control group (Shoji et al., 1992). High-frequency power ratio values were measured at 1 kHz intervals from 1 to 10 kHz for each vowel spectrum (Shoji et al., 1992). A statistically significant difference was found between the patient and control groups for all frequencies from 4 kHz to 10 kHz (Shoji et al., 1992). No statistically significant difference was found between genders for values at any of the frequencies (Shoji et al., 1992). It was proposed that the high-frequency power ratio found at 6 kHz performed the best in distinguishing between the patient and control groups and that an amplitude of –30 dB was the upper limit of normal voice (Shoji et al., 1992). Specifically, all but one of the patients showed a high-frequency power ratio value at 6 kHz above –30 dB (Shoji et al., 1992). All but one of the normal controls had a high-frequency power ratio value at 6 kHz below –30 dB (Shoji et al., 1992). These findings suggest that breathy voice may be associated with a spectral change in the high frequency region.

Another feature of breathy voice that distinguishes it from normal voice is a greater rounding of the glottal waveform, which creates larger first harmonic amplitudes (Hillenbrand et al., 1994). Perceptual studies by Bickley (1982) and Klatt and Klatt (1990) found that the amplitude of the first harmonic (H1) relative to the second harmonic (H2) to be a good predictor of breathiness, whereas Hillenbrand et al. (1994) found it to be only moderately correlated with perceived breathiness. The correspondence of the measure of the prominence of the first harmonic and breathiness will be discussed in more detail in Section 1.2.3.2.3 (“Amplitude Difference between Harmonics One and Two”).

1.2.3.2 Acoustic Measures Employed in This Study

The focus of the following section is on the acoustic parameters used in this study. These include frequencies of the first two formants (i.e., F1 and F2), amplitude difference between the first two harmonics (i.e., H1-H2), singing power ratio (i.e., SPR), consonant-to-vowel energy ratio (i.e., CV energy ratio), and voice onset time (i.e., VOT).

1.2.3.2.1 Formants One and Two (F1 and F2)

Vowels are produced with vocal fold vibration but largely without the constriction of oral air flow by vocal tract articulators involved in consonant production (Reetz & Jongman, 2009). Vowel differentiation is determined by advancement and elevation of the tongue and rounding of the lips (Robb & Chen, 2008). The position of the highest point of the tongue along with lip rounding are used to describe vowels phonetically (Reetz & Jongman, 2009). The four vowels used in this study were: /i/ as in "reach", /ɐ/ as in "arch", referred to in this study as /a/; /ɔ/ in "long", an open /o/ referred to in this study as simply /o/; and /u/ as in "two". These vowels are single vowels that can be referred to as corner vowels because they represent the extreme points where the tongue can be positioned in both vertical and horizontal dimensions. The vowel /i/ would be described as high, front (tongue position), and neutral (lip rounding), the vowel /a/ as low, central, and neutral, /o/ as low, back, and rounded, and /u/ as high, central and rounded (Reetz & Jongman, 2009).

Vocal tract shape determines formant frequencies (Reetz & Jongman, 2009). Acoustic energy becomes bunched up in its path through the vocal tract into areas of greater intensity called formants (Titze, 2000). Formants result from resonance (Baart, 2010). The many harmonics created at the glottis by vocal fold vibration are filtered by the vocal tract (Titze, 2000). Formants are seen on spectrograms as horizontal bands of increased energy. Spectrograms are plots with frequency on the vertical axis, time on the horizontal axis, and amplitudes represented by shading (darker shading for increased amplitude). Formants are commonly measured on the linear predictive coding (LPC) spectrum (Reetz & Jongman, 2009). The LPC spectrum, with amplitude on the vertical axis and frequency on the horizontal axis, shows vocal tract effects without displaying vocal fold periodicity (Reetz & Jongman, 2009). Formants appear as peaks in the LPC spectrum.

Measurements of formant frequencies should be taken from the middle of a vowel, preferably where it is most stationary and least influenced by sounds before or after it (Baart, 2010). In this study, formant frequencies were measured in the LPC spectrum from the temporal midpoint of the vowels. This avoided the need to use different measuring criteria for different formant configurations such as rising and falling formants (Baart, 2010).

Formants are labelled from low to high frequency as F1, F2, and F3. Frontness and height are acoustically correlated with F1 and F2 frequencies respectively (Torre III & Barlow, 2009). The frequency of F1 is inversely related to vowel height, with F1 frequency increasing as tongue height decreases (Torre III & Barlow, 2009). Vowels of similar height have similar F1 frequencies (Reetz & Jongman, 2009). Frontness is related to F2, more specifically the difference between F2 and F1 frequencies (Reetz & Jongman, 2009). The frequency of F2 is lower for the back vowels and higher for the front vowels (Reetz & Jongman, 2009).

1.2.3.2.2 Vowel Space

Vowels can be represented on the acoustic vowel space, a plot with F1 frequencies on the horizontal axis and F2 on the vertical axis. The acoustic vowel space signifies the accuracy of vowel articulation, reflecting the gross motor ability of the tongue and jaw coordination (Liu et al., 2005). The “corner vowels” /i, u, a/ from the high/front, high/back and low/back regions of vowel space are commonly used to measure acoustic vowel space

(Liu et al., 2005). The corner vowels are considered maximally distinctive vowels produced at the most physiologically extreme and stable articulatory positions of the vowel space (Robb & Chen, 2008).

A study of the acoustic-phonetic correlates of speech intelligibility by Bradlow et al. (1996) using sentence material produced by normal adult talkers revealed that intelligibility was generally greater for talkers with larger vowel spaces. As previously mentioned, in examining the acoustic properties of clear speech produced at normal speaking rates, Krause and Braida (2004) found that the vowel spaces measured from clear speech were larger than those measured from conversational speech. Picheny (1986) compared clear and conversational speech and found that in conversational speech, more frequently than in clear speech, vowels were reduced to become more neutral, mid-central or schwa-like. Poor formation of vowels in the speech of the hearing impaired, as previously mentioned, has also been found to result in reduced vowel space and vowels with a neutral indistinct schwa sound (Monsen, 1978).

1.2.3.2.3 Amplitude Difference Between Harmonics One and Two (H1-H2)

The H1-H2 amplitude difference is a measure of the dB amplitude of the first harmonic relative to the second harmonic (Hillenbrand et al., 1994). The H1-H2 amplitude difference was identified in a study by Bickley (1982) as consistently showing breathy from clear vowels in two languages that use breathiness for phonetic contrast. The speech materials consisted of recordings of South African Bushman producing clear and breathy versions of the vowel /a/ in their native language !Xóõ and native Gujarati talkers speaking from a word list (Bickley, 1982). Spectral analysis revealed that for 9 out of the 10 samples of the !Xóõ breathy vowels, the amplitude of the first harmonic exceeded that of the second harmonic (Bickley, 1982). For all of the 80 Gujarati samples of breathy vowels, the amplitude of the first harmonic was greater than that of the second harmonic (Bickley, 1982). Two phonetically trained listeners, a native English speaker and a native Gujarati speaker, made judgements as to the degree of breathiness of the vowels (Bickley, 1982). The vowels judged most breathy were found to have the greatest amplitude difference (Bickley, 1982). A further perceptual experiment was carried out using a synthesized continuum of clear to breathy versions of /a/, /i/, and /o/ Gujarti vowels where only H1 and the degree of aspiration noise were adjusted (Bickley, 1982). They were combined with natural Gujarati consonants to make one syllable Gujarati words and presented to four native Gujarati talkers for identification (Bickley, 1982). Vowels with the highest first harmonic amplitudes were consistently identified as breathy (Bickley, 1982). Inverse filtering of several clear and breathy naturally produced Gujarati vowels provided a means to observe the glottal waveforms without the effects of sound radiation and vocal tract filtering (Bickley, 1982). The glottal waveforms of the clear vowels had slower opening than closing phases, abrupt closures, and closed phases lasting a third of a cycle (Bickley, 1982). The glottal waveforms of breathy vowels had less abrupt closures and shorter closed periods than the clear waveforms, suggesting the likelihood that the glottis at no point achieved complete closure (Bickley, 1982). Bickley (1982) hypothesized that enhanced H1 is the acoustic correlate of breathiness.

In their study of variations of voice quality among male and female voices, Klatt and Klatt (1990) investigated H1-H2 amplitude difference referred to as “the amplitude of the first harmonic” as an acoustic correlate of breathiness. Six male and 10 female normal speaking talkers produced reiterant speech samples in which normal syllables were replaced by /ʔɑ/ or /hɑ/ (Klatt & Klatt, 1990). The amplitude of the first harmonic relative to the second harmonic was measured at vowel midpoints to avoid the influence of adjacent consonants (Klatt & Klatt, 1990). There was a difference of about 5.7 dB between the averaged harmonic differences measured from male and female talkers; the female talkers having the highest first harmonic amplitudes (Klatt & Klatt, 1990). As previously mentioned, their study found that female voices were on average perceived to be more breathy than male voices (Klatt & Klatt, 1990). Out of 10 acoustic measures associated with breathiness the H1-H2 amplitude difference was one of only two measures that had a statistically significant correlation with perceptual judgements of breathiness when natural voice material was used (Klatt & Klatt, 1990). The other measure was the degree of aspiration noise seen in the F3 region (Klatt & Klatt, 1990).

The H1-H2 amplitude difference was one of the measures used in a study by Hillenbrand et al., (1994) designed to evaluate the ability of acoustic measures to predict breathiness ratings. Normal talkers produced normal, moderately breathy, and very breathy versions of sustained vowels which were rated by Speech-Language Pathology graduate students using a direct magnitude rating scale ranging from a large number for very breathy signals to a small number for no or little breathiness, scores that were rescaled to the same range of scores (Hillenbrand et al., 1994). The H1-H2 amplitude difference, referred to in the study as the first harmonic amplitude, was described as correlating moderately with breathiness ratings, yet of the six acoustic measures investigated in the study it correlated the least well with breathiness ratings (Hillenbrand et al., 1994).

In this study the H1-H2 amplitude difference measure was calculated by subtracting the H1 amplitude in dB from the H2 amplitude.

1.2.3.2.4 Singing Power Ratio (SPR)

The singing power ratio (SPR) is a measure of the relative dominance of energy in a particular frequency region where a singer's formant is typically located. It is defined as the power ratio of the greatest harmonics peak between 2–4 kHz and between 0–2 kHz as measured from the midpoint of vowels (Omori, Kacker, Carroll, Riley, & Blaugrund, 1996). The SPR describes the resonant tuning of the vocal tract, reflecting the vocal tract's suppression or amplification of the harmonics produced at the glottis (Watts, Barnes-Burroughs, Estis, & Blanton, 2006). In Watts et al, (2006) a lower SPR measure indicated increased energy in the higher harmonics analogous to a lower spectral tilt.

Singing power ratio is normally used for the evaluation of singing voices to detect the presence of the singer's formant, which is another way of representing the increased energy in the higher harmonics. The singer's formant is produced in particular by male opera and concert singers trained in Western classical style of singing (Sundberg, 1987). The singer's formant is referred to erroneously as an extra formant but in fact arises from a lowering of formant frequencies in sung vowels resulting in a clustering of the third, fourth, and fifth formants to the spectral region between 2–4 kHz (Sundberg, 1987). The smaller gaps between frequencies results in greater amplitude occurring in a frequency region which coincides with where the sound produced by an orchestra is relatively weak (Sundberg, 1987). The highest SPL of an orchestra is at about 500 Hz (Mendes, Rothman, Sapienza, & Brown, 2003). The singer's formant thus enables an opera singer's voice to be heard above orchestral accompaniment (Sundberg, 1987). The presence of the singer's formant is thought to enhance the positively perceived richness and ringing of a singer's voice and correlates with lower SPR measures (Lundy, Roy, Casiano, Xue, & Evans, 2000; Watts et al., 2006).

The SPR measure has been proposed as a means to qualitatively evaluate voice quality of singers (Omori et al., 1996). A ringing voice quality was found to correspond with the presence of the singer's formant (Omori et al., 1996). Omori et al. (1996) used SPR to compare sustained vowels sung by trained singers to those sung by participants without training and to compare the trained singers' sung vowels to spoken versions. The speech materials were judged perceptually by experienced voice teachers using two scales extending from thin to rich and dull to ringing (Omori et al., 1996). The SPR measure found to correlate significantly with perceptual ratings of ringing (Omori et al., 1996). It was found that SPR measures from sung samples from singers were significantly higher than sung samples from non-singers (Omori et al., 1996). It was also found that SPR measures were significantly higher in the singers sung samples than their spoken samples (Omori et al., 1996).

Different results were found in a study by (Lundy et al., 2000) comparing SPR values measured from sung and spoken vocal samples recorded from singing students from a university school of music. Students were required to speak a sustained /a/ and sing a sustained /a/ (Lundy et al., 2000). Comparison was made between sung and spoken voices and normative data sets. There was no statistically significant difference found between the sung and spoken SPR values nor was there any significant difference based on years of training (Lundy et al., 2000).

A study by Watts et al. (2006) investigated SPR measures attained from a pool of untrained singers classed into two groups, talented and non talented, by a panel of professional voice teachers. Singing power ratio was measured from vocalic segments of recorded samples of their singing (Watts et al., 2006). The voice samples with the greater energy in the higher harmonics were found to correspond to the singers judged to be talented (Watts et al., 2006).

To measure SPR a Fast Fourier transformation is performed on the spectrum and the highest intensity peaks in the 0–2 kHz and 2–4 kHz frequency bands are measured. Lundy (2000) calculated SPR by dividing the "singing power peak" the greatest peak in the 2–4 kHz band by the greatest peak in the 0–2 kHz band. Watts (2006) calculated SPR by subtracting the highest peak in the 0–2 kHz band from the highest peak in the 2–4 kHz band.

In this study the greatest peak measured in dB from the 2–4 kHz band was subtracted from the greatest peak in the 0–2 kHz band.

1.2.3.2.5 Consonant-to-Vowel (CV) Energy Ratio

It has been long established that consonant sounds are lower in intensity than vowel sounds (French & Steinberg, 1947). The consonant-vowel ratio (CV energy ratio) is defined by Picheny et al. (1986) as "the ratio of the power of a consonant to that of the nearest vowel in the same syllable". The CV energy ratio was identified as a feature that distinguished between talkers in an article by House, Williams, Heker, and Kryter, (1965) the aim of which was the development and evaluation of a speech intelligibility test. There were two talkers, AH and CW, both adult males, experienced in the recording of materials for listening tasks (House et al., 1965). The CVC words produced by AH, the talker whose words were identified less well, had CV energy ratios 2-4 dB poorer on average regardless of whether the consonant was in the initial or final position (House et al., 1965). The maximum root-mean-square (RMS) values of the vowel portion of the CVC words were the same for the two talkers (House et al., 1965).

The first article in the "Speaking Clearly for the Hard of Hearing" series of articles examining differences between clear and conversational speech found that speaking clearly in contrast to speaking conversationally provided a gain in intelligibility of 17 percentage points averaged across the three talkers in the study (Picheny, Durlach, & Braida, 1985). In relation to CV energy ratio the second article in the series examined acoustic characteristics that could account for the gain and found that RMS intensities for obstruent sounds (sounds that restrict airflow that include stop consonants, fricatives and affricatives), especially stop consonants, were greater in clear speech than in conversational speech (Picheny et al., 1986).

Studies by Gordon-Salant (1986, 1987) aimed to determine the effect of modifications to consonant duration and CV energy ratio on consonant recognition in noise. Nonsense CV syllables were presented at either 75 dB SPL or 95 dB SPL to young and elderly participants with normal hearing in the first study and in the second study to elderly participants with sensorineural hearing loss (Gordon-Salant, 1986, 1987). The syllables were presented with and without acoustic modifications; the modification to CV energy ratio consisting of a 10 dB increase of consonant energy relative to vowel energy. This modification resulted in significant improvements to consonant recognition (Gordon-Salant, 1986). In the second study, alteration to CV energy ratio resulted in an improvement in nonsense syllable scores (percentage correct) of 14% and a reduction in the frequency of major consonant confusions (Gordon-Salant, 1987). The improvement to consonant recognition was attributed to an improvement in the "signal (consonant)-to-noise ratio" in part due to a reduced backward masking of the consonant by the vowel (Gordon-Salant, 1986, 1987).

In a study to determine the effect of increasing the consonant intensity on consonant recognition (not in noise) (Kennedy, Levitt, Neuman, & Weiss, 1998) used nonsense vowel-consonant syllables incorporating combinations of three vowels paired with nine voiced and seven unvoiced consonants. Each combination had different CV energy ratios in their unprocessed state (Kennedy et al., 1998). Consonant energy was increased up to 24 dB in 3 dB steps, the level of increase depending on the listener's dynamic range (Kennedy et al., 1998). The listeners consisted of three groups of sensorineural hearing impaired adults with different hearing loss configurations, one group with flat audiograms, one with sloping audiograms, the final group with steeply sloped audiograms (Kennedy et al., 1998). Stimuli were presented with the vowel at the listener's most comfortable loudness level as determined for each listener (Kennedy et al., 1998). The consonant intensity was then increased in 3 dB steps to a maximum level of 24 dB (Kennedy et al., 1998). If the participant reported an uncomfortable loudness level (UCL) less than 24 dB, the maximum intensity increase was limited to 3 dB below the reported UCL (Kennedy et al., 1998). It was found that the gain in recognition scores due to increases in consonant intensity were influenced most heavily by consonant type, and to a lesser degree by vowel pairing (Kennedy et al., 1998). The largest gains in consonant recognition, equivalent to as many as 45.8 percentage points, were achieved with voiceless stops /t/ and /k/ (which feature in the present study) and strong voiceless fricatives /s/ and /sh/ (Kennedy et al., 1998). The smallest gains were found with weak fricatives /ð/ and /θ/ and nasal consonants (Kennedy et al., 1998). A few combinations suffered a decline in consonant recognition score resulting from increases to consonant intensity, and some combinations produced no improvement or a plateau or a peak followed with a decline in improvement with further increases (Kennedy et al., 1998). Some vowel consonant combinations achieved 100% consonant recognition scores without consonant intensity increase (Kennedy et al., 1998). Listeners audiogram type had only a small effect on consonant recognition scores (Kennedy et al., 1998). The study showed that substantial improvements in consonant intelligibility were achievable with adjustments to the CV ratio that catered to individual listeners' hearing capacities (Kennedy et al., 1998).

1.2.3.2.6 Voice Onset Time (VOT)

Stop consonants or plosives are produced after vocal tract articulators close together to briefly stop oral air flow (Reetz & Jongman, 2009). Voice onset time (VOT) measures the time lapse between the release of a stop consonant and the onset of vocal fold vibration for the following voiced phoneme (Torre III & Barlow, 2009). Whilst it might last only a few milliseconds, this interval contains important information as to plosive identity (Stouten & Van hamme, 2009). For example a shorter VOT will distinguish between /b/ in "buy" from

/p/ in “pie” (Torre III & Barlow, 2009). Voice onset time describes the control and timing of vocal tract articulators involved in producing stop consonants (Torre III & Barlow, 2009).

The stop consonants used in this study are /p/, /t/ and /k/ from the words “pot”, “people”, “two”, “take” and “colours”. They are all word initial unvoiced stop consonants. Voiceless stop consonants have a relatively long gap between the release of the stop and the onset of voicing (Auzou et al., 2000). In contrast, voiced stop consonants (/b, d, g/) have a shorter gap between stop release and vocal fold vibration (Auzou et al., 2000). The stop consonants in this study differ in terms of place of articulation, the following vowel, and sentence position. “People” initiates a sentence, whereas “pot”, “two”, “take” and “colours” have words preceding them. The closure for /p/ is bilabial (lips together), /t/ is tongue to alveolar ridge, and /k/ is tongue to roof of mouth (velar).

Analysis of the acoustics of normal speech in a study by Volaitis and Miller (1992) confirmed findings of previous studies that place of articulation influenced VOT to the effect that a move in place of articulation from labial to alveolar to velar resulted in a lengthening of VOT. The study also confirmed that this occurs across a range of speaking rates to both voiced and voiceless stop consonants (Volaitis & Miller, 1992). VOT varies with the rate of speech (Torre III & Barlow, 2009). Increases to speaking rate result in a reduced VOT for voiceless stop consonants and has little effect on voiced stop consonants (Baum & Ryan, 1993).

A study by Allen, Miller and DeSteno (2003) investigating differences in VOT between talkers found that talkers differ in VOT values and that the major predictor of difference was differences in speaking rate among talkers. They found that when speaking rate was controlled for, VOT continued to differ among talkers to a lesser degree (Allen et al., 2003). In a follow-up study, Theodore, Miller and DeSteno (2009) showed that the size of increase to VOT for each stop consonant for a given change in speaking rate varied between talkers. Across the speech of an individual talker, speaking rate influences VOT for one voiceless stop in the same way as it does for other voiceless stops (Theodore et al., 2009).

Investigating the acoustic characteristics of clear relative to conversational speech, Picheny et al. (1986) found that VOT of unvoiced stop consonants was increased in clear speech whereas VOT of voiced stop consonants was increased for the clear speech of only one of the three talkers in the study. Fricatives, nasals, and semivowels were also found to have durational increases in clear speech (Picheny et al., 1986).

Monsen (1978) compared acoustic measures and intelligibility scores from the speech of 37 adolescents with severe-to-profound hearing loss. Voice onset times were measured from readings of list words in order to measure participants ability to produce distinctions between voiced and unvoiced stop consonants (Monsen, 1978). Voicing for the voiced stops was measured from the release of the occlusion, not prior to it, because even in careful speech, voiced stops are often devoiced and voicing occurring during closure is subject to error due to being at times difficult to discern from spectrograms (Monsen, 1978). Averaged VOTs of voiced stops were subtracted from those of voiceless counterparts (sharing place of articulation) (Monsen, 1978). The study found that distinction between the stop consonants correlated highly with intelligibility (Monsen, 1978).

Torre III and Barlow (2009) in an investigation into changes to acoustic properties of speech related to aging, found that older men had shorter VOTs than older women, which is consistent with previous findings of men having shorter VOT s than women. According to Torre III and Barlow (2009) there are many studies that have found that women produce longer VOTs than men regardless of age. Torre III and Barlow (2009) also found that older men had shorter VOTs than younger men. They also found more variability in the speech of older talkers than same gender younger talkers (Torre III & Barlow, 2009).

Koenig (2000) listed factors that impinge on VOT in addition to timing and control of articulators as the degree of vocal fold closure; glottal channel shape; sub-, trans-, and supra-glottal pressure levels; and vocal fold tissue differences including stiffness and compliance. Knowledge that these factors vary within normal populations in accordance with age and sex prompted Koenig's (2000) investigation as to whether stop consonant VOTs alongside voicing characteristics of the glottal fricative /h/ varied predictably with age and sex. Data was gathered from normal speaking adults and 5-year-old children of both genders (Koenig, 2000). It was found that VOT acquisition in children, often not consistent with those of adults until puberty is reached, rely on development of voicing as well as interarticular timing control (Koenig, 2000).

1.3 Research Outline

This study includes two parts. In the first part of the study, a selection of acoustic measures, including F1 and F2 formant frequencies, the H1-H2 amplitude difference, SPR, CV energy ratio, and VOT, were derived from speech samples recorded from adult voice patients showing complete and incomplete glottal closure during phonation. Statistical analysis was conducted on the measures to gauge the effect of the independent variable, conditions of glottal closure (complete vs. incomplete glottal closure) on these measures. In the second part of this study, speech samples were selected on the basis of H1-H2 amplitude difference level and were presented for listener judgment in two perceptual tasks: vowel discrimination and clarity discrimination. The results were analysed to determine changes in perceptual ratings as a function of the H1-H2 amplitude difference levels.

1.3.1 Statement of the Problem

Speech intelligibility is an aspect of human communication that relates to the comprehension of sounds, words, and phrases. Disorders of voice can greatly impinge upon a person's ability to communicate effectively. The human voice can be evaluated using subjective and objective measures. The precise relationship between specific voice characteristics and the intelligibility of speech is unclear. To identify key features for maintaining good speech and voice quality, the relationship between the acoustic and perceptual measures of voice is in need of clarification.

1.3.2 Aims of the Study

The purpose of this study is to delineate the relationship between voice quality, in particular, breathiness, and speech intelligibility. The aims were to identify, via acoustic analysis, the voice characteristics of voice patients with and without complete glottal closure, and to determine the impact of voice changes related to breathiness on vowel intelligibility and the perception of vowel clarity.

1.3.3 Hypotheses

It is speculated that voice quality would have an impact on speech intelligibility. This speculation will be tested acoustically and perceptually. Acoustically, the main hypotheses to be tested in the Stage One analysis of the study are as follows:

In general, it was hypothesized that the acoustic measures used in this study, including F1 and F2 frequencies, H1-H2 amplitude difference, SPR, CV energy ratio, and VOT would differentiate between disordered voice samples produced by voice patients with complete glottal closure from those produced by voice patients with incomplete glottal closure.

Specifically, it was hypothesised that in comparison with voice produced by voice patients with complete glottal closure, voices produced by those with incomplete glottal closure would show:

changes in F1 and F2 frequencies resulting in decreased acoustic vowel spaces, which is associated with reduced speech intelligibility;

a more dominant H1 resulting in lower values of H1-H2 amplitude difference, which is suggestive of a more breathy voice;

a lower SPR value indicating decreased voice projection power;

a reduction in CV energy ratio, which is indicative of a reduced consonant intensity relative to vowel intensity;

and longer VOT due to a delay in voice initiation.

In the second part of the study, the speech stimuli used in the perceptual tasks were selected according to levels of H1-H2 amplitude difference, a measure chosen because it had been shown in previous studies to correlate with breathiness. It was hypothesized that the presence of more breathiness as evidenced by a greater dominance of H1 would result in a decrease in intelligibility of vowels and a reduction in the perception of clarity.

2 Method

This study involved two stages of data collection and analysis to determine how changes of the acoustic characteristics of pathological voice may impact on speech intelligibility. Stage one involved acoustic analysis of samples of pathological voice collected from patients with complete and incomplete glottal closure. Stage two consisted of a perceptual study conducted to verify whether the acoustic features pertaining to breathiness were related to the perception of vowel identification and clarity.

2.1 Stage One: Acoustic Study

The first stage required the measurement and analysis of acoustic parameters conducted on voice samples. The voice samples included segments taken from digitized voice recordings of voice patients producing running speech and sustained vowels.

2.1.1 Voice Recordings

The voice recordings used in this study were previously acoustically recorded from voice patients seen in the voice clinic in the Otolaryngology Department at the Christchurch Hospital. All recordings were made in a sound booth used for audiometric testing in the hospital. The voices varied in the degree of pathology they exhibited. The voice samples were classified according to whether the patients achieved complete or incomplete glottal closure under videolaryngostroboscopic examination. Digitized voice files of 26 voice patients, including 13 cases associated with complete and 13 cases with incomplete glottal closures were retrieved for acoustic analysis. The "complete glottal closure" group included 7 males (aged from 32 to 65 years; Mean = 46.7, SD = 12.2) and 6 females (aged from 29 to 54; Mean = 40.0, SD = 9.9) and the "incomplete glottal closure" group included 7 males (aged from 24 to 81; Mean = 48.4, SD = 20.3) and 6 females (aged from 43 to 68;

Mean = 55.3, SD = 10.8). The recordings consisted of sustained vowels along with readings of the first six sentences in The Rainbow Passage (Fairbank, 1960). Words with similar syllabic and phonetic structure, each with a different vowel, were selected from the sentences (see Appendix 1).

2.1.2 Instrumentation

The acoustic recording system consisted of a headset microphone (AKG C420, Austria) and a mixer (Eurorack MX602A, Behringer) used as microphone preamplifier. The output of the mixer was connected to a 12-bit A/D converter (National Instrument DAQCard-AI-16E-4, USA) via a SCB-68 68-pin shielded connector box. The connector box contained a filter for the acoustic signals to be low-passed at 20 KHz. The A/D converter was housed by a laptop computer (Compaq 650 MHz Pentium 4, Taiwan) for direct digitization. Time-frequency analysis software TF32 (Milenkovic, 2001) was used to perform analysis of the acoustic signals.

2.1.3 Acoustic Measures

The acoustic parameters measured from the mid portion of selected vowels included the frequencies of the first two formants, the amplitude difference between the first two harmonics, and singing power ratio. Maximum RMS values were measured from consonants and vowels for the calculation of the CV energy ratio. Voice onset time was measured from words initiated with a stop consonant.

2.1.3.1 Vowels

An approximately 50 ms-long segment was sectioned out from each of the words "reach" (for the vowel /i/), "arch" (/ɑ/ referred to in this study as /a/), "long" (/ɔ/ referred to in this study as /o/), and "two" (/u/), which were embedded in the third or sixth sentences of the Rainbow Passage. Portions of 500 ms duration (starting proximal to the initial ramped onset) were sectioned from recordings of sustained vowels /a/ and /i/. Frequencies of the first two formants (based on linear predictive coding spectrum), amplitudes of the first two harmonics (long-time average spectrum), and amplitudes of the highest spectral peaks identified in the 0–2 and 2–4 kHz frequency bands (long-time average spectrum), were measured from the "sustained vowels" and "sentence-embedded vowels". Voice onset time and CV energy ratio were measured from "sentence-embedded vowels".

2.1.3.1.1 F1 and F2

The frequencies of the first two formants of vowels (/a, i, o, u/) were measured. On the spectrogram display, a cursor was positioned at the temporal mid-point of the vowel. An LPC (Linear Predictive Coding) spectrum was displayed in a separate window. The "LP" (LPC spectrum) and "Pre-emphasis" parameters were selected. Frequencies of F1 and F2 were selected for measuring by manually aligning the cursor with the selected spectral peaks.

2.1.3.1.2 Vowel Space

Frequencies of F1 and F2 extracted from vowels (/a, i, u/) that were later selected for use in the perceptual study were used to calculate vowel space areas. First and second formant frequencies measured from "sentence-embedded vowels" /a, i, u/ of individual talkers, along with a set of "sentence-embedded vowels" /a, i, u/ from pooled stimuli (not specific to individual talkers, were plotted in an F1-F2 space to form vowel space triangles. The area of each triangle was calculated with the following formula quoted from Liu et al. (2005):

Vowel triangle area = ABS{[F1i*(F2a-F2u)+F1a*(F2u-F2i)+F1u*(F2i-F2a)]/2}, where "ABS" is absolute value, "F1i" symbolises the F1 value of vowel/i/, "F2a" symbolises the F2 value for the vowel /a/,… and so on.

2.1.3.1.3 H1-H2

The amplitudes of the first two harmonics were measured from vowels (/a, i, o, u/). On the spectrogram display, cursors were aligned to select portions of the vowels (approximately 50 ms of the vowels embedded in words and the first 500 ms of the steady portion of the sustained vowels). An LTA (long-time average) spectrum for the selected vowel segment was displayed in a separate window. The "H" parameter (Hanning window) was selected. The amplitudes of H1 and H2 were determined by aligning the cursors with the corresponding harmonics to generate an automatic readout value.

2.1.3.1.4 SPR

The SPR was measured from vowels (/a, i, o, u/). On the spectrogram display, cursors were aligned to select portions of the vowels (approximately 50 ms of the vowels embedded in words and the first 500 ms of the steady portion of the sustained vowels). An LTA spectrum was displayed in a separate window. The "Pre-emphasis" and "H" (Hanning Window) parameters were selected. The amplitudes of the highest peaks that fell between 0 and 2 kHz, and between 2 and 4 kHz were determined by aligning the cursors with the corresponding spectral peaks to generate an automatic readout value. An increase in the energy of the spectrum in the 2–4 kHz frequency region will yield a larger SPR value.

2.1.3.2 VOT

Voice onset time of stop consonants /p/, /t/, and /k/ were measured from the words "pot", "people", "two", "take", and "colours". Displays of the time waveform and spectrogram were used to judge cursor positioning at the beginning and end of the initial consonant to give a time measurement.

2.1.3.3 CV Energy Ratio

Maximum RMS values were measured for consonants and vowels for the calculation of the CV energy ratio for words “arch”, “reach” “long”, and “two”. End points were selected from the RMS trace display. One cursor was positioned to capture the maximum RMS for the consonant and the other cursor was positioned for the vowel. The zoom function was used to help judge cursor placement.

2.1.4 Statistical Analysis of Acoustic Measures

A series of two-way mixed design between-within subjects Analysis of Variances (ANOVAs) were conducted on the acoustic measures for each of the “sentence-embedded vowels”, “sustained vowels”, and selected words (in the case of VOT and CV energy ratio) separately. The statistical tests conducted on the F1, F2, H1-H2, and SPR measures were to determine whether there was a glottal closure effect (complete versus incomplete glottal closure), vowel effect (/a/, i/, /o/, and /u/), or glottal closure by vowel interaction effect on these measures. For VOT and CV energy ratio, two-way (glottal closure by word) mixed design between-within subjects ANOVA were conducted to determine whether there was a glottal closure effect (complete versus incomplete glottal closure) or word effect (“pot”, “people”, “two”, “take”, and “colours” for VOT; “reach”, “arch”, “long”, and “two” for CV energy ratio) on these measures.

Because only four of the voice patients had voice samples recorded from both before and after treatment, all the samples were treated as individuals to maximize the sample size.

Data was divided into male and female to account for gender differences. Statistical analysis was carried out using Sigma Stat version 3.5. The significance level was set at 0.05.

2.1.5 Reliability

To determine the reliability of the acoustic measurement, more than 20% (96 out of 400 "male and female sentence-embedded vowels", 24 out of 96 selected words for CV energy ratio measures, 25 out of 115 selected words for VOT measures, and 56 out of 244 "male and female sustained vowels") were reanalysed by the same experimenter several months after the original measurements were made. Results from a series of Pearson Product Moment Correlation procedures revealed a relatively high measure-remeasure reliability for all acoustic measures, including F1 ($r = 0.848$), F2 ($r = 0.560$), SPR ($r = 0.624$) and H1-H2 amplitude difference ($r = 0.787$) from "sentence-embedded vowels", F1 ($r = 1.000$), F2 ($r = 0.985$), SPR ($r = 0.999$), and H1-H2 amplitude difference ($r = 0.983$) from "sustained vowels", and CV energy ratio ($r = 0.997$) and VOT ($r = 0.990$) from selected words.

2.2 Stage Two: Perceptual Study

The second stage of this study involved the participation of listeners in perceptual tasks. Judgements were obtained that would determine the intelligibility of speech samples that were selected on the basis of the level of H1-H2 amplitude difference.

2.2.1 Participants and Participants' Task

A convenience quota sampling method was used for participant recruitment. Recruitment was via personal request, e-mail request, and posters displayed around the University of Canterbury campus. The participants were 10 female and 10 male adult native English speakers aged between 19 and 34 years with normal speech, language, and hearing history. Participants were required to have normal hearing on the day of testing as defined as thresholds no greater than 20 dB HL at octave frequencies 0.5 to 4 kHz. Screening audiometry was conducted immediately before the listening tasks. Pure tone thresholds were measured down to 20 dB from an initial presentation at 30 dB using the ASHA recommended modified Hughson-Westlake Procedure as described by Harrell (2002). Screened thresholds were obtained at octave intervals from 0.5 to 8 kHz. All of the listeners had normal hearing, defined as thresholds no greater than 20 dB HL, at octave intervals from 0.5 to 4 kHz. Two of the listeners had slight monaural hearing loss at 8 kHz with thresholds at 25 dB HL. One of the listeners had mild binaural hearing loss at 8 kHz, with thresholds at 40 dB HL in one ear and 30 dB HL in the other ear.

During the experiment, participants were asked to listen to sets of stimuli and performed forced choice tasks based on their perception of the stimuli. There were two tasks, "vowel identification" task and "clarity discrimination" task. In the "vowel identification" task, the participant was asked to listen to one vowel token at a time and identify from a list of five vowels which vowel was heard. In the "clarity discrimination" task, the participant was asked to listen to one pair of vowel tokens in each trial and select the token that was perceived as "clearer" in the pair. At the end of the experiment, participants were given petrol vouchers to the value of $20 to compensate for their participation.

2.2.2 Stimuli

The acoustic stimuli that were presented to listeners in the perceptual study were selected from the digitized voice files used for the acoustic analysis. The stimuli included both "sentence-embedded vowels" and "sustained vowels". The 50 ms segments sectioned from sentence readings ("sentence-embedded vowels"), including vowels /a, i, o, u/ segmented from the words, "reach", "arch", "long", and "two", were used in the "vowel identification" task. The 500 ms tokens taken from sustained vowels /a, i/ ("sustained vowels") were used in the "clarity discrimination" task. The selection of the stimuli was based on H1-H2 amplitude difference measures to allow for inclusion of a range of voice with different degrees of breathiness because the H1-H2 measure has been identified in the literature review as being correlated with breathiness ratings.

The stimuli were grouped by gender and vowel so that for each gender there were four groups of "sentence-embedded" stimuli and two groups of the "sustained vowel" stimuli. The range of H2-H1 amplitude difference measures for the stimuli within each group was determined. Each range was divided evenly into four divisions. The midpoints of each of the divisions (four midpoints per range) marked off five intervals: three equal sized intervals between the midpoints and two half-sized intervals outside the midpoints to the ends of each range. The intervals served as levels one through to five, spanning the range of H1-H2 amplitude difference measures for each range, with level one having the lowest values, which indicates H1 dominance being the strongest (i.e. presumably most breathy). Up to two stimuli were selected as samples from each of the five levels for each range. Selection from the two end levels included the outermost measure of the range as 'sample one', and its nearest neighbour within the level as 'sample two'. For Levels two to four, the most central value was selected as "sample one" and its nearest neighbour as "sample two". An uneven spread of values within the ranges meant that selection for some of the levels was lacking or limited to one sample.

The selections formed two sample groups each for the "vowel identification" and "clarity discrimination" tasks. The four ranges, each with five levels of H1-H2 amplitude difference values for male and female "sentence-embedded vowels", are shown in Appendices 4 and 5 respectively. The two ranges, each with five levels of H1-H2 amplitude difference values for male and female "sustained vowels", are shown in Appendices 2 and 3 respectively.

Stimuli were randomized within the selections using online randomizer software Research Randomizer (Urbaniak, Plous, & Lestik, 2010). The stimuli for the "clarity discrimination" task, within the two selections grouped by gender and vowel were paired so that each stimulus was paired with every other within group stimuli, in both positions pair-wise. The pairs were then randomized. For the "vowel identification" task, genders and vowels were randomized together.

A total of 20 (5 H1-H2 levels X 2 vowels X 2 genders) vowel tokens were used in each task. The twenty 50 ms vowel segments taken from sentence readings and twenty 500 ms segments taken from sustained vowel recordings were produced by the same talkers. As vowel segments taken from embedded vowels were used in the "vowel identification" task while those taken from sustained vowels were used in the "clarity discrimination" task, a correlation procedure was conducted on the H1-H2 measures for the two types of speech stimuli used in the perceptual study. As shown in Figure 1, the Pearson Product Moment Correlation resulted in a correlation coefficient ($r = 0.639$) indicating that there is a moderate positive correlation between the two types of speech stimuli. The two types of stimuli are significantly related ($P < 0.002$).

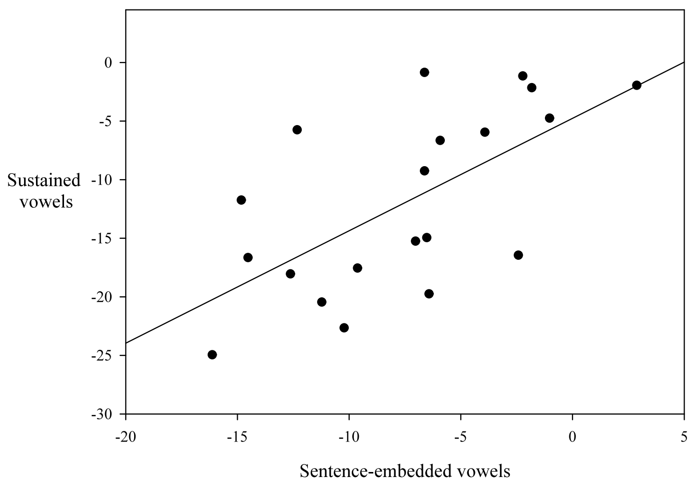


Figure 1. Correlation between "sustained vowels" and "sentence-embedded vowels".

2.2.3 Instrumentation

Adobe Audition software was used to normalize the intensity of the sound samples used as tokens in the perceptual tasks. The listening tasks were carried out in a sound booth at the University of Canterbury Communication Disorders Department. A Grason Stadler GSI 61 audiometer with TDH50 supra aural headphones was used to screen hearing. A locally developed computer algorithm written in C++ was installed in a desktop computer equipped with a high-quality sound card to present the stimuli and record responses.

2.2.4 Procedure

Participants were seated in the sound booth. Stimuli were presented to the participant via headphones and the participant was asked to perform the participant's task. For the "vowel identification" task, listeners selected the vowel, represented on the screen in front of them, which best approximated their perception of what they heard (see Appendix 6). They could repeat the sounds before selection. For the discrimination task, the listeners indicated which they perceived to be the "clearer" of two tokens presented in pairs (see Appendix 7). As with the identification task, the listeners could opt to listen to the pairs of vowels as often as they chose. Instruction was provided on the interface screen of the programme. An additional verbal instruction was given for the "vowel identification" task that the listener should focus on the word examples not the symbol preceding the word examples. Listeners were advised not to "agonise" over their decisions to discourage them from over thinking and applying their own internal criteria. The hearing screen and set of listening tasks took about an hour for each participant. Consent forms were signed prior to the listening tasks and petrol vouchers were dispensed once finished.

2.2.5 Statistical Analysis of Perceptual Measures

A series of one-way ANOVAs were conducted on averaged listener responses to the male and female "sentence-embedded vowels" and "sustained vowels" presented separately in the "vowel identification" and "clarity discrimination" tasks. These were carried out to determine whether there was an H1-H2 level effect and whether such effect, if present, followed the expected trend of Level 1 stimuli (measured as having the lowest level of H1-H2 amplitude difference and thus the greatest H1 dominance) receiving the lowest percentage correct and lowest percentage clear scores with scores increasing with increasing levels of H1-H2 amplitude difference. Post-hoc pair-wise multiple comparison procedures using the Holm-Sidak method were performed when a significant effect was detected. The significance level was set at 0.05. Sigma Stat version 3.5 (Systat Software, Inc., USA) was used for all statistical analysis.

3 Results

Results from the statistical analysis of the acoustic measures are described in Section 3.1 ("Acoustic Measures") and results of the averaged participant judgements in the "vowel identification" and "clarity discrimination" tasks in Section 3.2 ("Perceptual Measures").

3.1 Acoustic Measures

Results from a series of two-way mixed between-within subjects ANOVAs performed on the acoustic measures obtained from sentence-embedded and sustained vowel recordings were shown in Tables 1 and 2 separately for measures of F1 and F2 frequencies, H1-H2, and SPR and results for measures obtained from sentence-embedded words in Table 3 for CV energy ratio and VOT.

3.1.1 F1

The ANOVA results for the F1 acoustic measure obtained from "male and female sentence-embedded vowels" separately showed that there was a significant vowel effect for both male and female patients but no significant glottal closure effect nor glottal closure by vowel interaction effect for either gender (see Table 1). The absence of significant difference between glottal closure conditions for both genders is shown in Figure 2. As for vowel effect, it is shown in Figure 3 that for both genders, all vowels were significantly different from one another apart from /i/ and /u/ between which there was no significant difference. As shown in Figure 3, for both males and females, the high vowels (/i, u/) exhibited lower F1 frequency than the low vowels (/a, o/) as expected.

Similarly, the ANOVA results for the F1 acoustic measure obtained from "female and male sustained vowels" separately, showed that there was a significant vowel effect for the male and for the female patients, no significant glottal closure effect nor glottal closure by vowel interaction effect for either gender (see Table 2). The absence of significant difference between glottal closure conditions for both genders is shown in Figure 4. As for vowel effect, it is shown in Figure 5 that the vowel /a/ had a significantly higher F1 than the vowel /i/ for both genders as expected.

Table 1. Glottal closure and vowel effect on acoustic measures from "sentence-embedded vowels".
Two-way (glottal closure by vowel) mixed between-within subjects ANOVA results calculated separately for acoustic measures obtained from "male and female sentence-embedded vowels".

		Glottal Closure Effect		Vowel Effect		Glottal Closure by Vowel	
F1	Male	$F(1,30) = 0.334$	$p = 0.571$	$F(3,30) = 19.987$	$p < 0.001$*	$F(3,30) = 0.565$	$p = 0.642$
	Female	$F(1,33) = 2.990$	$p = 0.112$	$F(3,33) = 32.377$	$p < 0.001$*	$F(3,33) = 0.396$	$p = 0.757$
F2	Male	$F(1,30) = 0.293$	$p = 0.118$	$F(3,30) = 9.212$	$p < 0.001$*	$F(3,30) = 2.220$	$p = 0.106$
	Female	$F(1,33) = 0.690$	$p = 0.424$	$F(3,33) = 21.217$	$p < 0.001$*	$F(3,33) = 0.675$	$p = 0.573$
H1-H2	Male	$F(1,30) = 17.090$	$p = 0.002$*	$F(3,30) = 8.761$	$p < 0.001$*	$F(3,30) = 1.537$	$p = 0.225$
	Female	$F(1,33) = 1.689$	$p = 0.220$	$F(3,33) = 12.898$	$p < 0.001$*	$F(3,33) = 2.430$	$p = 0.083$
SPR	Male	$F(1,30) = 0.009$	$p = 1.000$	$F(3,30) = 6.049$	$p = 0.002$*	$F(3,30) = 0.855$	$p = 0.475$
	Female	$F(1,33) = 1.967$	$p = 0.188$	$F(3,33) = 4.544$	$p = 0.009$*	$F(3,33) = 0.453$	$p = 0.717$

* Significant at 0.05 level.
Vowel effect is related to vowels /a, i, o, u/.

Table 2. Glottal closure and vowel effect on acoustic measures from "sustained vowels".
Two-way (glottal closure by vowel) mixed between-within subjects ANOVA results calculated separately for acoustic measures obtained from "male and female sustained vowels".

		Glottal Closure Effect		Vowel Effect		Glottal Closure by Vowel	
F1	Male	$F(1,14) = 2.986$	$p = 0.106$	$F(1,14) = 103.869$	$p < 0.001$*	$F(1,14) = 2.424$	$p = 0.142$
	Female	$F(1,12) = 0.187$	$p = 0.673$	$F(1,12) = 49.354$	$p < 0.001$*	$F(1,12) = 0.686$	$p = 0.424$
F2	Male	$F(1,14) = 7.151$	$p = 0.018$*	$F(1,14) = 14.296$	$p = 0.002$*	$F(1,14) = 7.292$	$p = 0.0017$*
	Female	$F(1,12) = 0.481$	$p = 0.501$	$F(1,12) = 10.381$	$p = 0.007$*	$F(1,12) = 3.449$	$p = 0.088$
H1-H2	Male	$F(1,14) = 2.973$	$p = 0.107$	$F(1,14) = 15.227$	$p = 0.002$*	$F(1,14) = 1.767$	$p = 0.205$
	Female	$F(1,12) = 2.248$	$p = 0.160$	$F(1,12) = 12.055$	$p = 0.005$*	$F(1,12) = 1.749$	$p = 0.211$
SPR	Male	$F(1,14) = 0.734$	$p = 0.406$	$F(1,14) = 25.549$	$p < 0.001$*	$F(1,14) = 11.942$	$p = 0.004$*
	Female	$F(1,12) = 0.002$	$p = 0.961$	$F(1,12) = 4.512$	$p = 0.055$	$F(1,12) = 1.668$	$p = 0.221$

* Significant at 0.05 level.
Vowel effect is related to vowels /i/ and /a/.

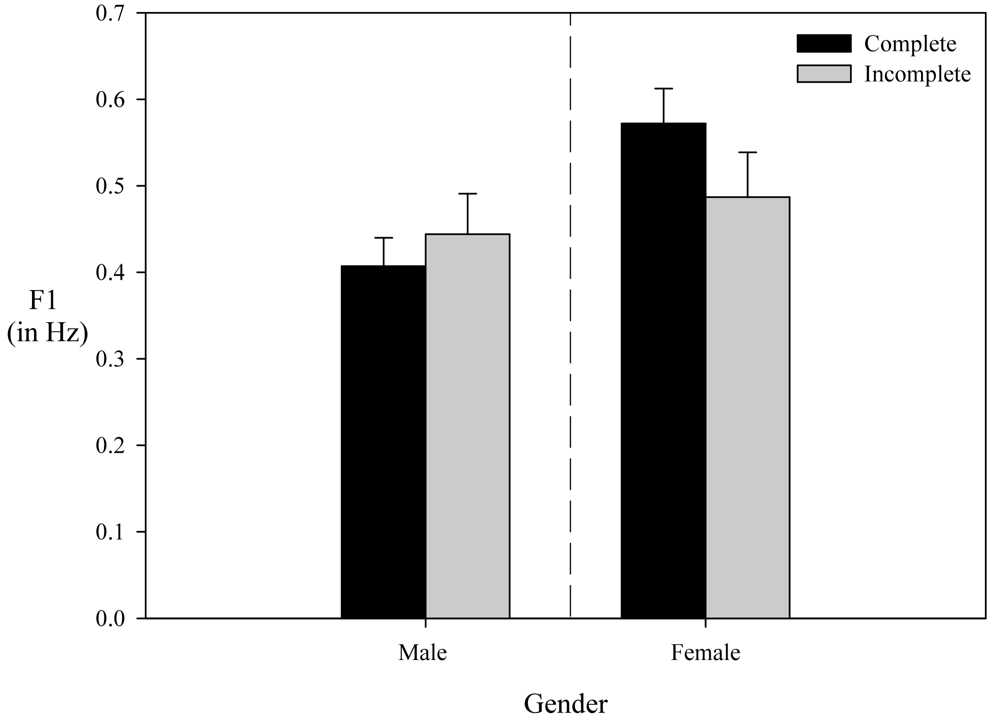


Figure 2. Glottal closure effect on F1 measured from “sentence-embedded vowels”.

Means and standard error of means for the F1 acoustic measures obtained from “male and female sentence-embedded vowels” for each glottal closure (complete vs. incomplete) in the male and female groups separately [male complete and incomplete (n = 24), female complete (n = 28) and incomplete (n = 24)]. Significantly different pairs in each data set are marked with an asterisk (“*”).

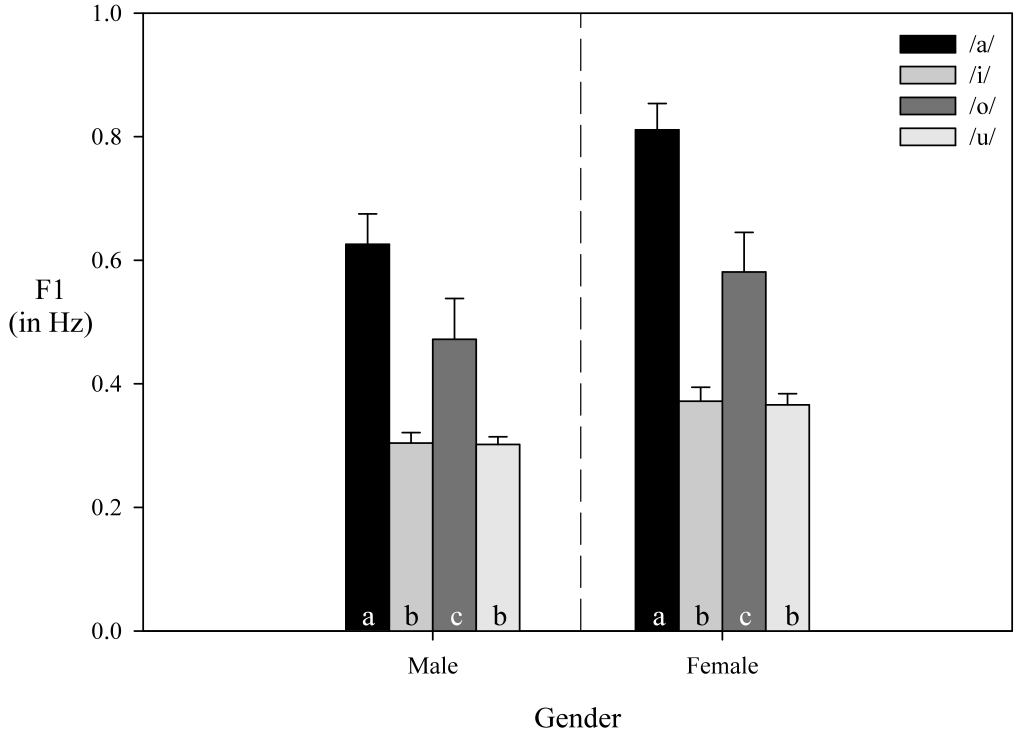


Figure 3. Vowel effect on F1 measured from "sentence-embedded vowels".

Means and standard error of means for the F1 measures obtained from "male and female sentence-embedded vowels" for each vowel (/a/, /i/, /o/, and /u/) in the male and female groups separately [male (n = 12), female (n = 13)]. Significantly different pairs in each data set are marked with different letters.

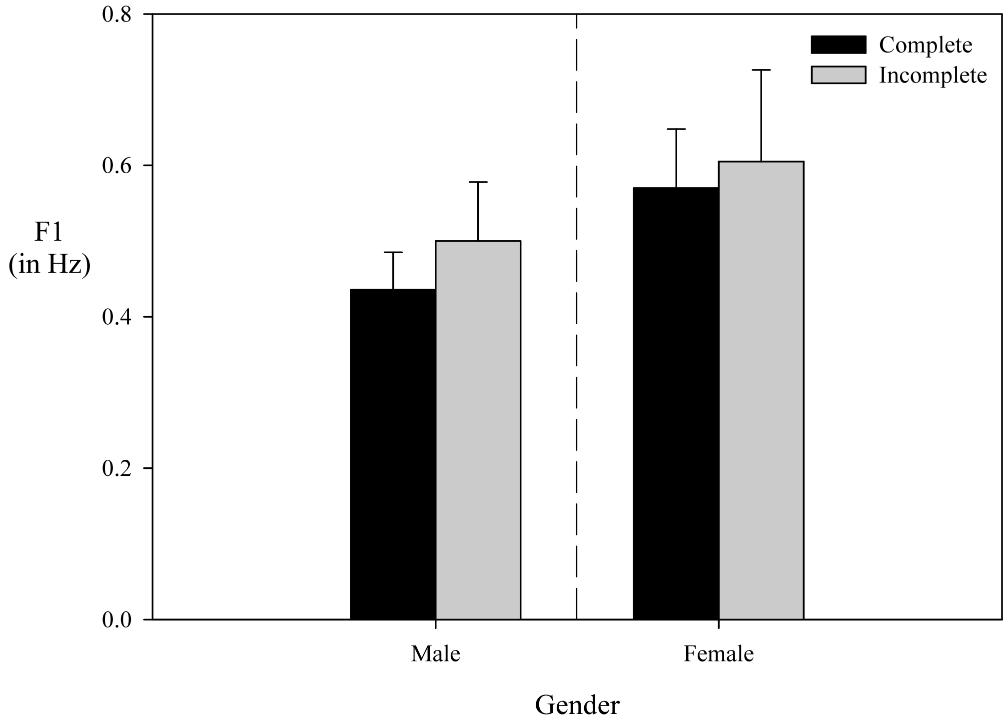


Figure 4. Glottal closure effect on F1 measured from "sustained vowels".

Means and standard error of means for the F1 measures obtained from "male and female sustained vowels" for each glottal closure condition (complete vs. incomplete) in male and female groups separately [male complete (n = 18) and incomplete (n = 14), female complete and incomplete (n = 14)]. Significantly different pairs in each data set are marked with an asterisk ("*").

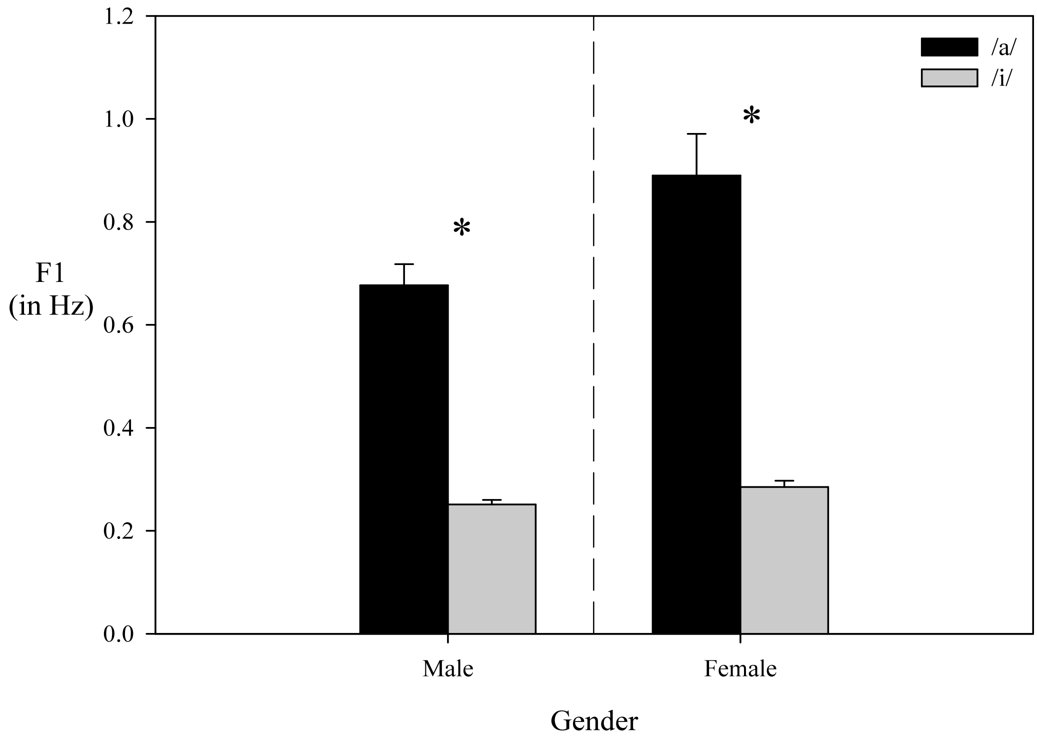


Figure 5. Vowel effect on F1 measured from "sustained vowels".

Means and standard error of means for the F1 measures obtained from "male and female sustained vowels" for the two vowels (/a/ vs. /i/) in the male and female groups separately [male (n = 16), female (n = 14)]. Significantly different pairs in each data set are marked with an asterisk ("*").

3.1.2 F2

The ANOVA results for the F2 acoustic measure obtained from "male and female sentence-embedded vowels" separately showed that there was a significant vowel effect for both the male and female patients (see Table 1). There was no significant glottal closure effect, nor glottal closure by vowel interaction effect for either gender (see Table 1). The absence of significant difference between glottal closure conditions for both genders is shown in Figure. As for vowel effect, it is shown in Figure 7 that for both genders the vowel /i/ exhibited a significantly higher F2 than the other three vowels, /a/, /o/, and /u/ but there was no significant difference between the two low vowels, /a/ and /o/. For female patients, /u/ had a significantly higher F2 than the two low vowels, /a/ and /o/.

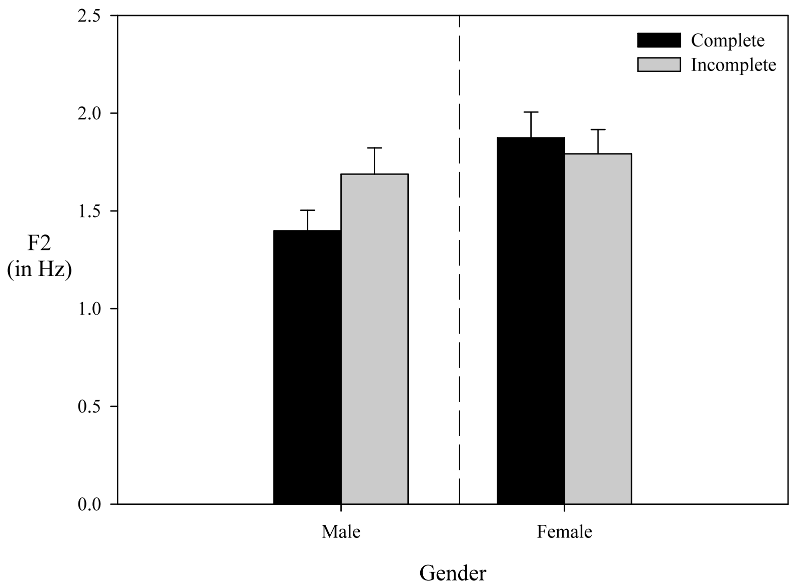


Figure 6. Glottal closure effect on F2 measured from "sentence-embedded vowels". Mean and standard error of means for the F2 measures obtained from "male and female sentence-embedded vowels" for each glottal closure condition (complete vs. incomplete) in the male and female groups separately [male complete and incomplete (n = 24); female complete (n = 28), incomplete (n = 24)]. Significantly different pairs in each data set are marked with an asterisk ("*").

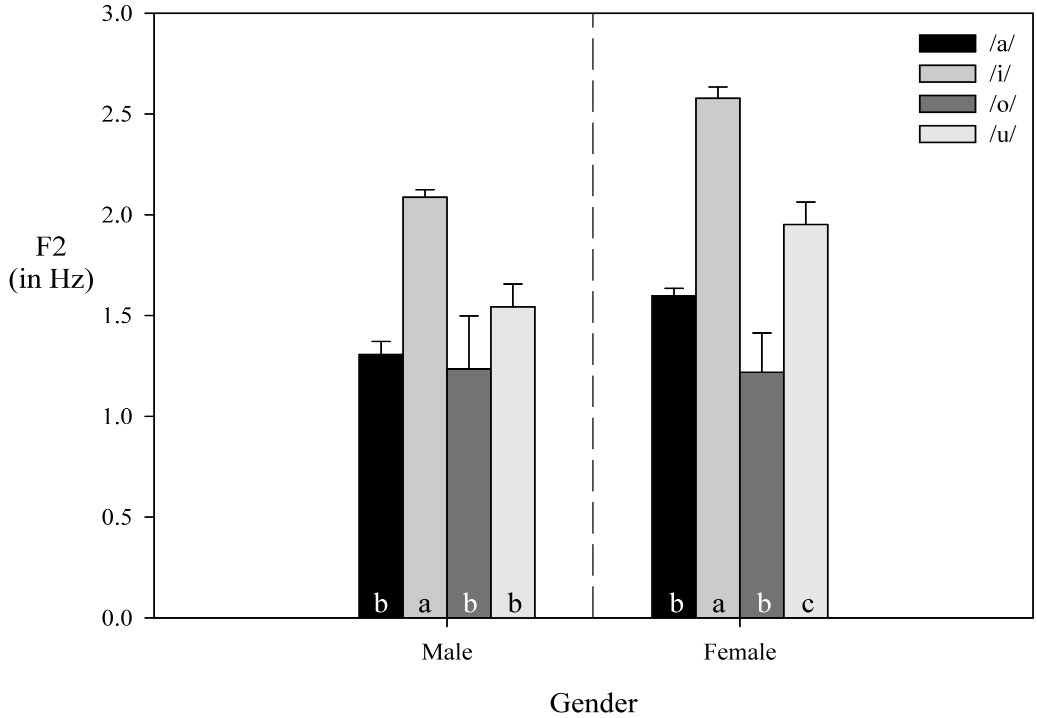


Figure 7. Vowel effect on F2 measured from "sentence-embedded vowels".

Means and standard error of means for the F2 measures obtained from "male and female sentence-embedded vowels" for each vowel (/a/, /i/, /o/, and /u/) in the male and female groups separately [male (n = 12); female (n = 13)]. Significantly different pairs in each data set are marked with different letters.

The ANOVA results for the F2 measure obtained from "male and female sustained vowels" separately, showed that there was a significant vowel effect, glottal closure effect, and glottal closure by vowel interaction effect for the male patients but only a significant vowel effect for the female patients (see Table 2). For both genders, as shown in Figure , the group with incomplete glottal closure tends to show a higher F2 than the group with complete glottal closure although the difference is significant only in the male data. As for vowel effect, it is shown in Figure 9 that, for both genders, the vowel /i/ had a significantly higher F2 than /a/ as expected. Since the male data also shows a significant glottal closure by vowel interaction effect, Figure 10 is presented to show the interaction effect in the "male sustained vowels" data. As shown in Figure 10, although /i/ tends to have a higher F2 than /a/ as expected, the difference is only significant for the group with complete glottal closure, suggesting that the normal maintenance of the difference in F2 between /a/ and /i/ may be lost in voices associated with incomplete glottal closure.

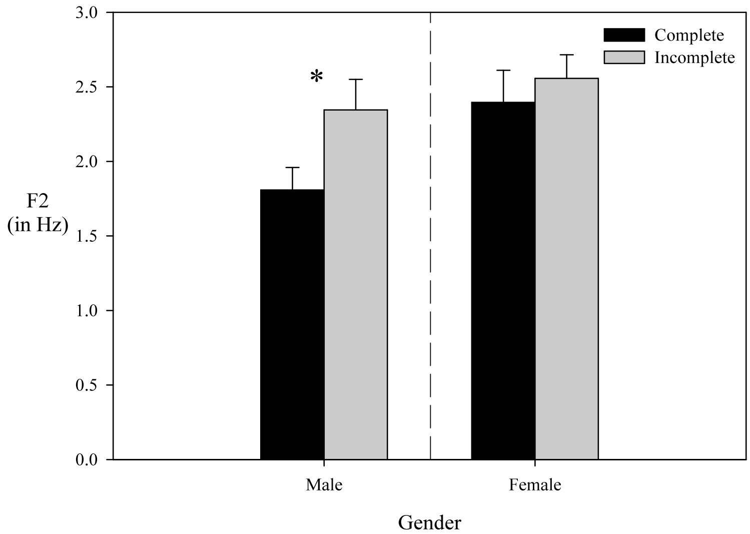


Figure 8. Glottal closure effect on F2 measured from "sustained vowels".
Means and standard error of means for the F2 measures obtained from "male and female sustained vowels" for each glottal closure condition (complete vs. incomplete) in the male and female groups separately [male complete (n = 18), incomplete (n = 14); female complete and incomplete (n = 14)]. Significantly different pairs in each data set are marked with an asterisk ("*").

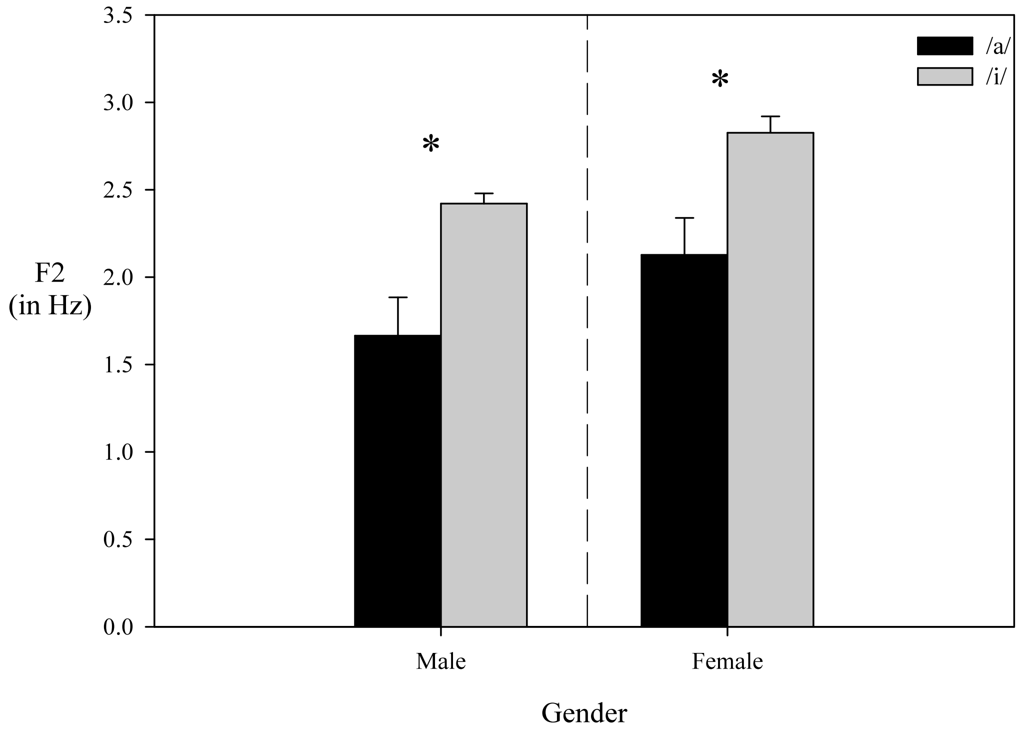


Figure 9. Vowel effect on F2 measured from "sustained vowels".

Means and standard error or means for the F2 measures obtained from "male and female sustained vowels" for each vowel (/a/ vs. /i/) in male and female groups separately [male (n = 16); female (n = 14)]. Significantly different pairs in each data set are marked with an asterisk ("*").

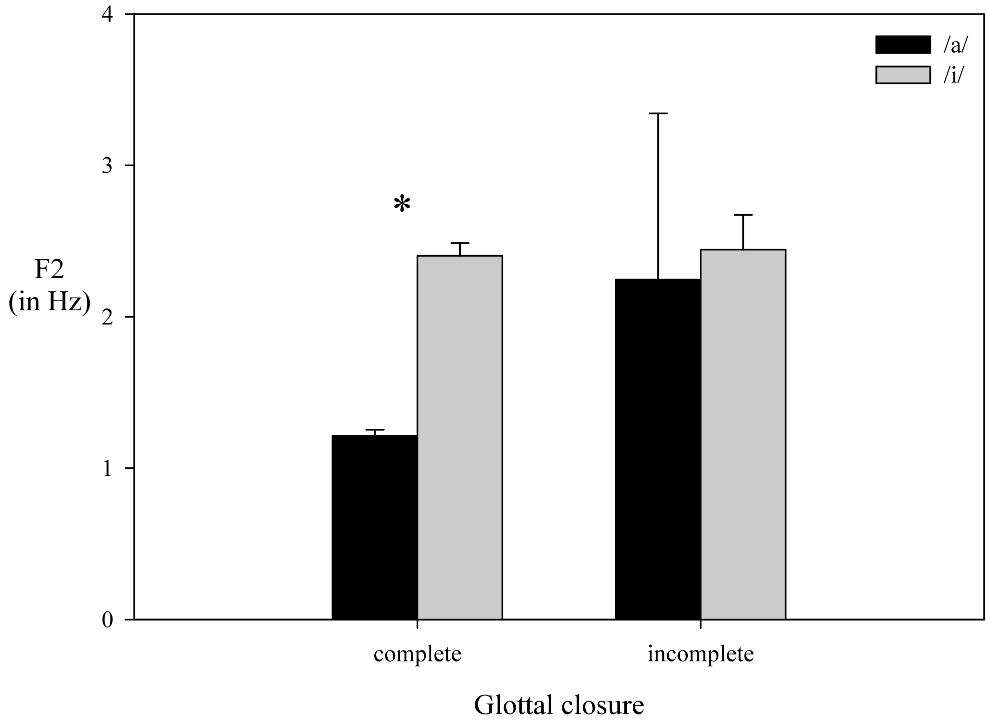


Figure 10. Interaction effect on F2 of vowel within glottal closure from "male sustained vowels".

Means and standard error of means for the F2 measures obtained from "male sustained vowels" for each glottal closure condition (complete vs. incomplete) separately [complete (n = 9), incomplete (n = 7)]. Significantly different pairs in each data set are marked with an asterisk ("*").

3.1.3 H1-H2

The ANOVA results for the H1-H2 amplitude difference acoustic measure obtained from "male and female sentence-embedded vowels" separately showed that there was significant glottal closure effect and vowel effect for the male patients and only a significant vowel effect for the female patients (see Table 1). There was no significant glottal closure by vowel interaction effect for either gender (see Table 1). It is shown in Figure that there was a significant difference between glottal closure conditions for male patients but not for female patients although in both genders, the group with incomplete glottal closure shows a significantly lower average H1-H2 score (i.e., greater H1 prominence and thus more breathy) than the group with complete glottal closure. As for vowel effect, for both genders, the vowel /a/ has a significantly lower H1-H2 (more breathy) than the vowel /i/ (see Figure 12). For female patients, the vowel /a/ also has a significantly lower H1-H2 than /u/.

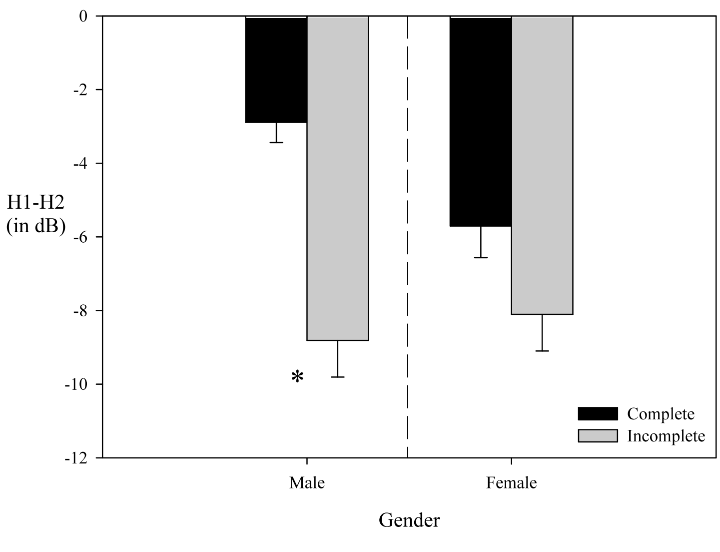


Figure 11. Glottal closure effect on H1-H2 measured from "sentence-embedded vowels".
Means and standard error of means for the H1-H2 (amplitude difference between the first two harmonics) measures obtained from "male and female sentence-embedded vowels" for each glottal closure condition (complete vs. incomplete) in the male and female groups separately [male complete and incomplete (n = 24), female complete (n = 28) and incomplete (n = 24)]. Significantly different pairs in each data set are marked with an asterisk ("*").

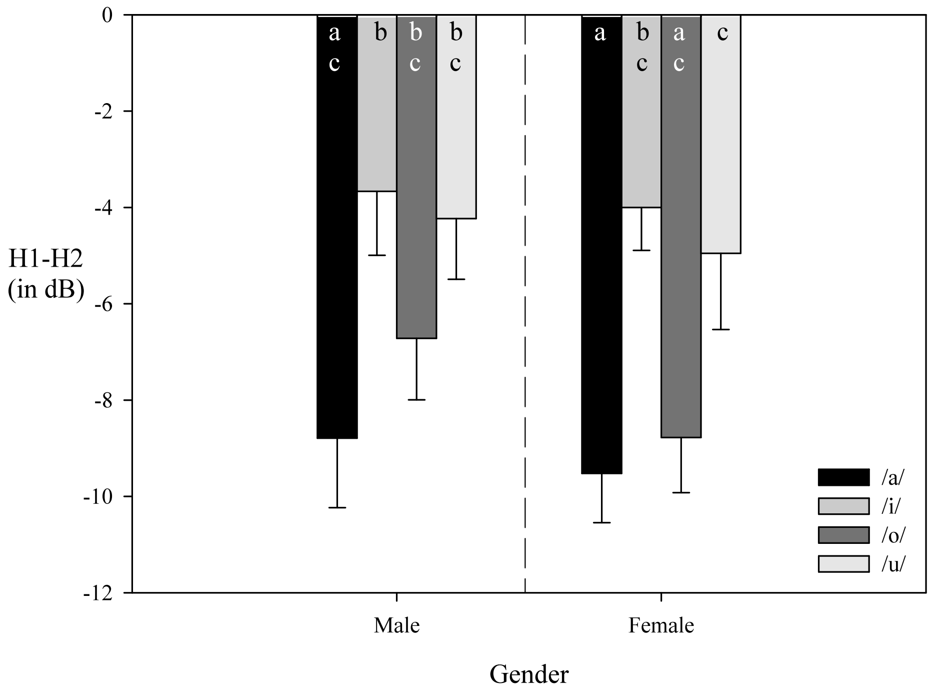


Figure 12. Vowel effect on H1-H2 measured from "sentence-embedded vowels".

Means and standard error of means for the H1-H2 (amplitude difference between the first two harmonics) measures obtained from "male and female sentence-embedded vowels" for each vowel (/a/, /i/, /o/, and /u/) for the male and female groups separately [male (n = 12), female (n = 13)]. Significantly different pairs in each data set are marked with different letters.

The ANOVA results for the H1-H2 amplitude difference acoustic measure obtained from "male and female sustained vowels" separately, showed that there was a vowel effect for patients of both genders, but no significant glottal closure effect nor glottal closure by vowel interaction effect for either gender (see Table 2). There lack of significant difference between glottal closure conditions for either gender is shown in Figure . It can be observed from Figure 13, however, the group with incomplete glottal closure tends to show a lower average H1-H2 value than the group with complete glottal closure. As for the vowel effect, it is shown in Figure 14 that the vowels /a/ and /i/ for both genders were significantly different from each other, with /a/ showing a lower average H1-H2 value than /i/ in the male data but a higher average H1-H2 value than /i/ in the female data.

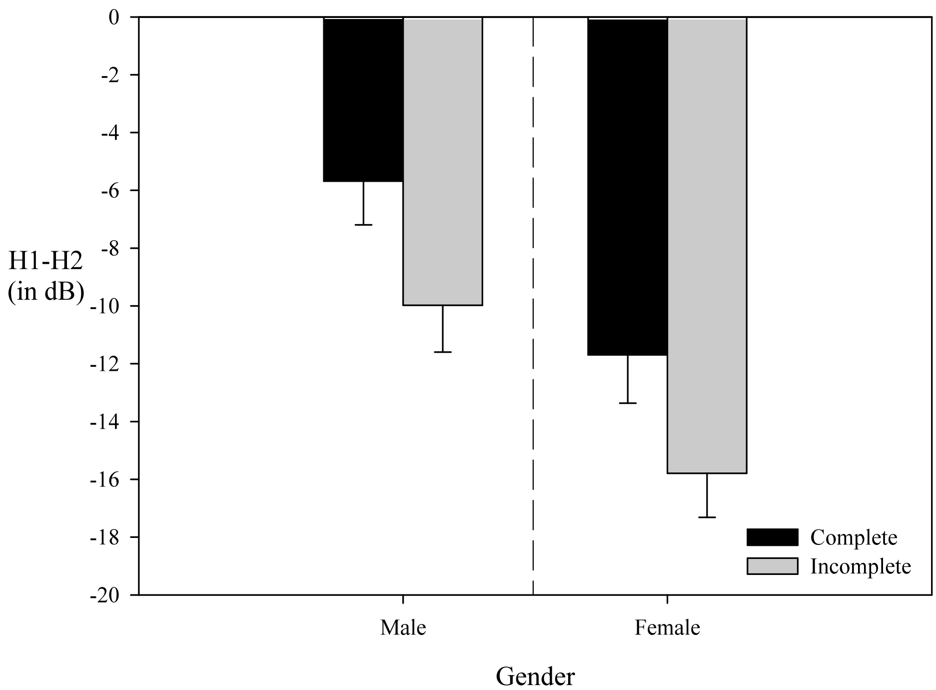


Figure 13. Glottal closure effect on H1-H2 measured from "sustained vowels".
Means and standard error of means for the H1-H2 (amplitude difference between the first two harmonics) measures obtained from "male and female sustained vowels" for each glottal closure condition (complete vs. incomplete) in the male and female groups separately [male complete (n = 18) and incomplete (n = 14), female complete and incomplete (n = 14)]. Significantly different pairs in each data set are marked with an asterisk ("*").

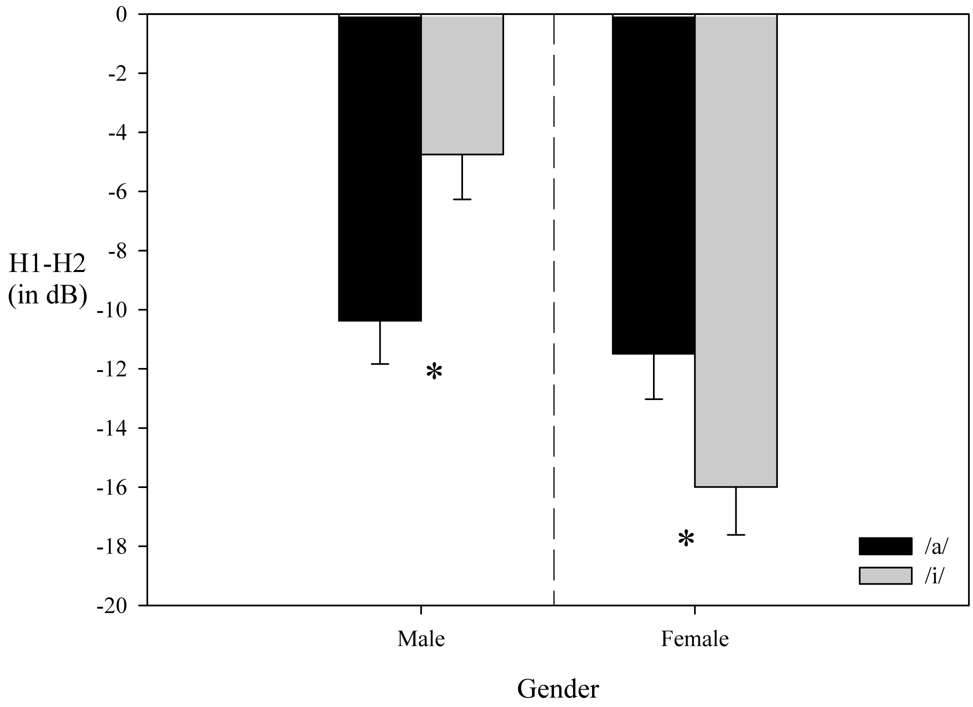


Figure 14. Vowel effect on H1-H2 measured from "sustained vowels".

Means and standard error of means for the H1-H2 (amplitude difference between the first two harmonics) measures obtained from "male and female sustained vowels" for each vowel (/a/ vs. /i/) in the male and female groups separately [male (n = 16), female (n = 14)]. Significantly different pairs in each data set are marked with an asterisk ("*").

3.1.4 SPR

The ANOVA results for the SPR acoustic measure obtained from "male and female sentence-embedded vowels" separately showed that there was a significant vowel effect for both genders (see Table 1). There was no significant glottal closure effect, nor glottal closure by vowel interaction effect for either gender (see Table 1). The absence of significant difference between glottal closure conditions for both genders is shown in Figure . As for the vowel effect, it is shown in Figure 16 that, for both genders, the vowel /i/ had a significantly higher average SPR value (suggestive of a stronger voice projection power) than /o/ and the vowel /a/ also had a significantly higher average SPR value than /o/.

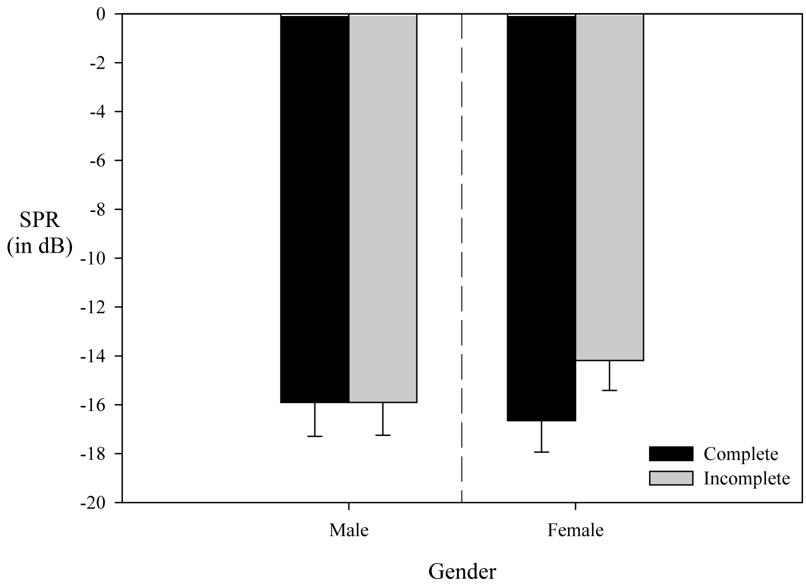


Figure 15. Glottal closure effect on SPR measured from "sentence-embedded vowels". Means and standard error of means for the SPR measures obtained from "male and female sentence-embedded vowels" for each glottal closure condition (complete vs. incomplete) in the male and female groups separately [male complete and incomplete (n = 24), female complete (n = 28), incomplete (n = 24)]. Significantly different pairs in each data set are marked with an asterisk ("*").

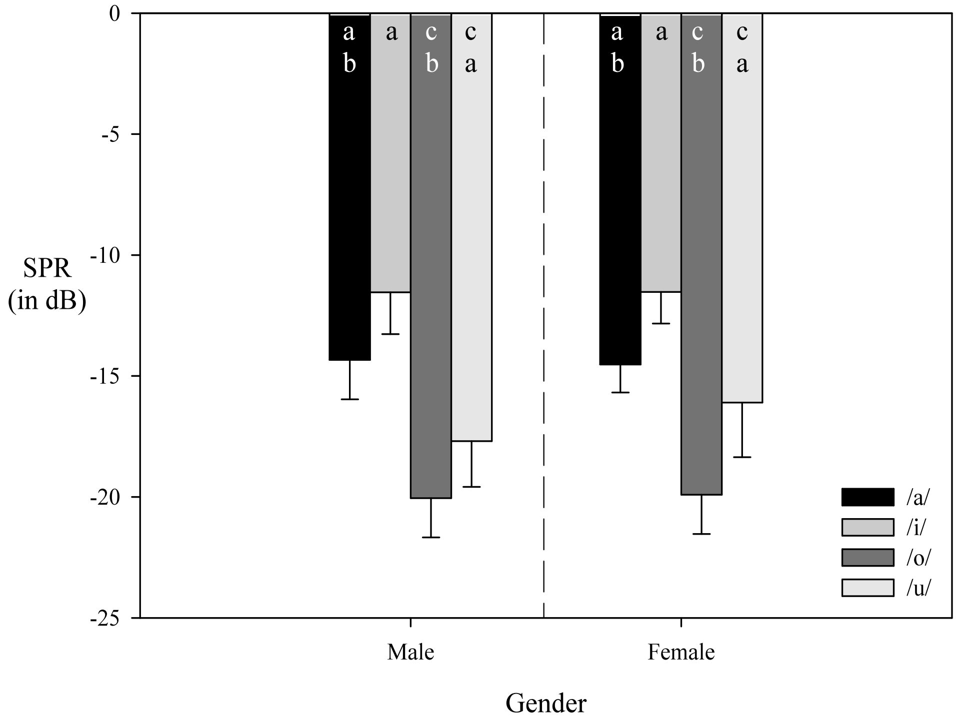


Figure 16. Vowel effect on SPR measured from "sentence-embedded vowels".

Means and standard error or means for the SPR measures obtained from "male and female sentence-embedded vowels" for each vowel (/a/, /i/, /o/, and /u/) in male and female groups separately [male (n = 12), female (n = 13)]. Significantly different pairs in each data set are marked with different letters.

The ANOVA results for the SPR acoustic measure obtained from “male and female sustained vowels” separately showed that there was a significant vowel effect and glottal closure by vowel interaction effect for the male patients (see Table 2). There was no significant glottal closure effect for either gender and no glottal closure by vowel interaction effect for the female patients (see Table 2). The absence of significant difference between glottal closure conditions for both genders is shown in Figure . It is shown in Figure 18 that the vowels /a/ and /i/ for males alone were significantly different from each other. As for the glottal closure by vowel interaction effect, there was a significant difference between the vowels in the incomplete glottal closure condition but not in the complete glottal closure condition (see Figure 19).

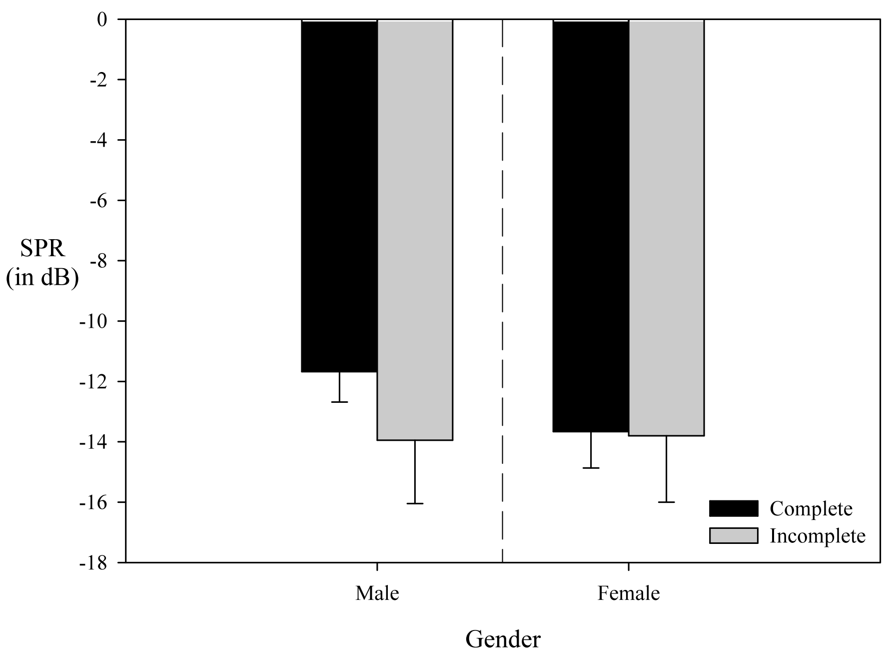


Figure 17. Glottal closure effect on SPR measured from “sustained vowels”.
Means and standard error or means for the SPR measures obtained from “male and female sustained vowels” for each glottal closure condition (complete vs. incomplete) in male and female groups separately [male complete (n = 18), incomplete (n = 14), female complete and incomplete (n = 14)]. Significantly different pairs in each data set are marked with an asterisk (“*”).

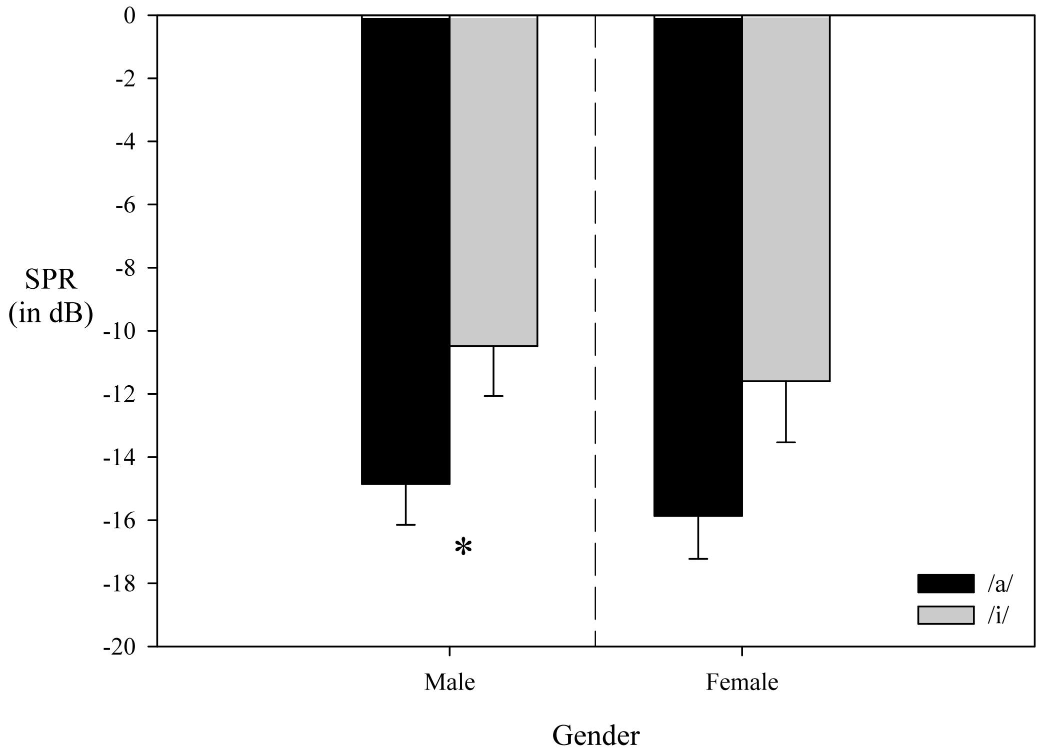


Figure 18. Vowel effect on SPR measured from "sustained vowels".

Means and standard error of means for the SPR measures obtained from "male and female sustained vowels" for each vowel (/a/ vs. /i/) in male and female groups separately [male (n = 16), female (n = 14)]. Significantly different pairs in each data set are marked with an asterisk ("*").

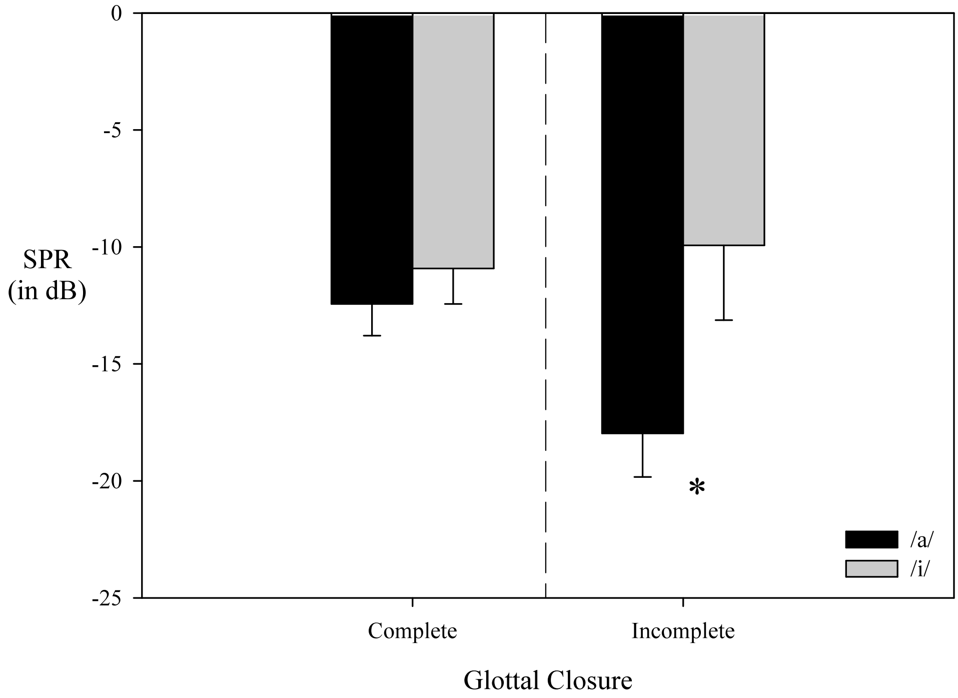


Figure 19. Glottal closure by vowel interaction effect on SPR for "male sustained vowels".

Means and standard error of means for the SPR measures obtained from "male sustained vowels" for each vowel (/a/ vs. /i/) in the complete and incomplete glottal closure conditions separately [complete (n = 9), incomplete (n = 7)]. Significantly different pairs in each data set are marked with an asterisk ("*").

3.1.5 CV Energy Ratio

The ANOVA results for the CV energy ratio measure obtained from the male and female groups separately showed that there was a significant word effect for the male patients alone (see Table 3). There was no significant glottal closure effect, nor glottal closure by word effect for either gender (see Table 3). The absence of a significant difference between glottal closure conditions for both genders is shown in Figure 20. As shown in Figure 20, the incomplete glottal closure condition tends to show a higher CV energy ratio than the complete glottal closure. However, due to the large within-group variation, the statistical test failed to reveal a significant glottal closure condition effect. For male patients, as shown in Figure 21, the word "arch" which contains the vowel /a/ shows a significantly higher CV energy ratio than all the other words ("reach", "long", and "two), which contain the vowels /i/, /o/, and /u/. For males, no significant difference was found amongst the words "reach", "long", and "two." There was no significant difference between words for female patients (see Figure 21).

Table 3. Glottal closure and vowel effect on CV energy ratio and VOT.
Two-way (glottal closure by vowel) mixed between-within subject ANOVA results calculated separately for acoustic measures obtained from selected words.

		Glottal Closure Effect		Word Effect		Glottal Closure by Word	
CV energy ratio	Male	$F(1,27) = 2.178$	$p = 0.174$	$F(3,27) = 6.430$	$p = 0.002$*	$F(3,27) = 0.496$	$p = 0.688$
	Female	$F(1,24) = 1.199$	$p = 0.305$	$F(3,24) = 1.401$	$p = 0.267$	$F(3,24) = 0.319$	$p = 0.812$
VOT	Male	$F(1,32) = 1.263$	$p = 0.294$	$F(4,32) = 2.492$	$p = 0.063$	$F(4,32) = 0.904$	$p = 0.473$
	Female	$F(1,44) = 0.951$	$p = 0.350$	$F(4,44) = 21.011$	$p < 0.001$*	$F(4,44) = 1.965$	$p = 0.116$

* Significant at 0.05 level.
Word effect is related to the CV energy ratio measure taken from “reach”, “long”, “arch” and “two” and the VOT measure taken from “pot”, “people”, “two”, “take” and “colours”.

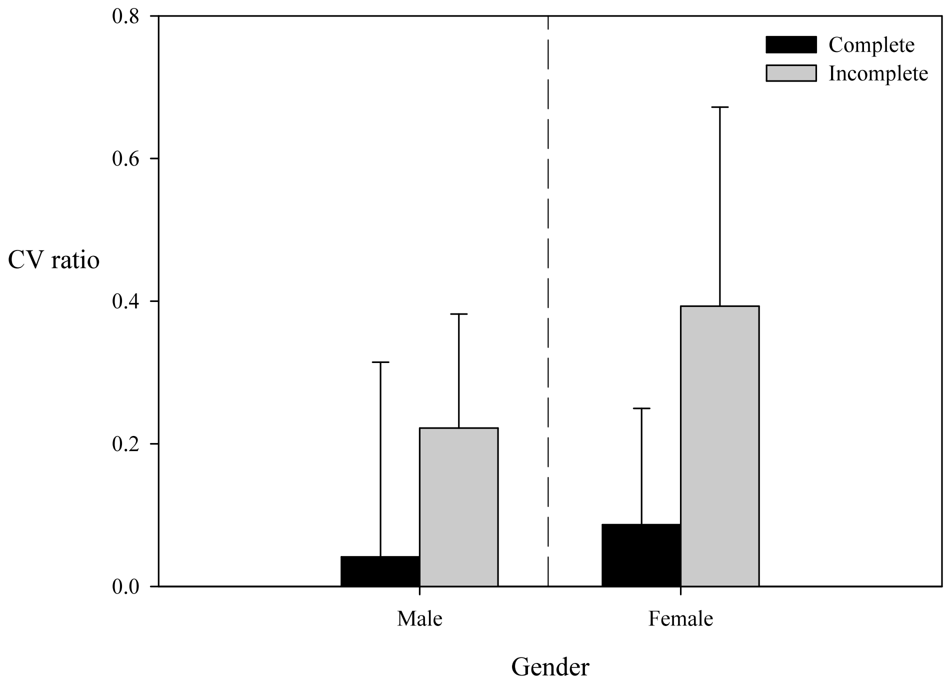


Figure 20. Glottal closure effect on CV energy ratio measured from selected words. Means and standard error of means for the CV energy ratio measures obtained from selected words for each glottal closure condition (complete vs. incomplete) in male and female groups separately [male complete (n = 24) and incomplete (n = 20), female complete (n = 16) and incomplete (n = 24)]. Significantly different pairs in each data set are marked with an asterisk ("*").

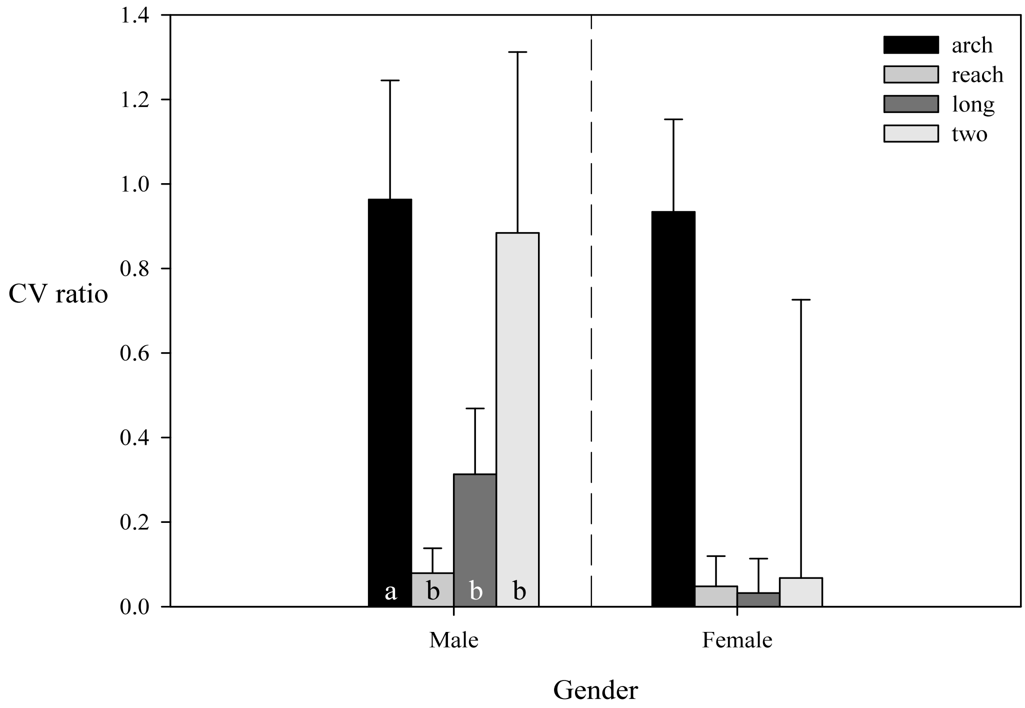


Figure 21. Word effect on CV energy ratio measured from selected words.

Means and standard error of means for the CV energy ratio measures obtained from selected words ("arch", "reach", "long", and "two") embedded with each vowel (/a/, /i/, /o/, /u/) in male and female groups separately [male (n = 11), female (n = 10)]. Significantly different pairs in each data set are marked with different letters.

3.1.6 VOT

The ANOVA results for the VOT measure obtained from male and female selected words separately, showed that there was a significant word effect for the female patients alone (see Table 3). There was no significant glottal closure effect, nor glottal closure by word effect for either gender (see Table 3). The absence of significant difference between glottal closure conditions for both genders is shown in Figure . The absence of significant difference between words for male patients is shown in Figure 23. It is shown in Figure 23 that, for female patients, there was significant change in VOT for the word “pot” in comparison with “people”, “take” and “colours”, for the word “two” in comparison with “people” and “take”, and for the word “colours” in comparison with “people” and “take” (see Figure 23). There was no significant change of VOT for the word “two” in comparison with “pot” and “colours”, nor for the word “people” in comparison with “take” (see Figure 23). In general, it appears that VOT differs significantly between consonants, with /k/ showing the longest VOT, followed in order by /t/ and /p/.

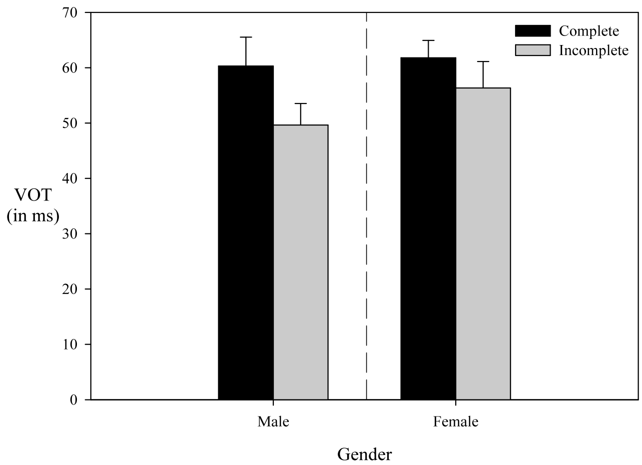


Figure 22. Glottal closure effect on VOT measured from selected words.
Means and standard error of means for the VOT measures obtained from selected words for each glottal closure condition (complete vs. incomplete) in the male and female groups separately [male complete and incomplete (n = 25), female complete (n = 35) and incomplete (n = 30)]. Significantly different pairs in each data set are marked with an asterisk (“*”).

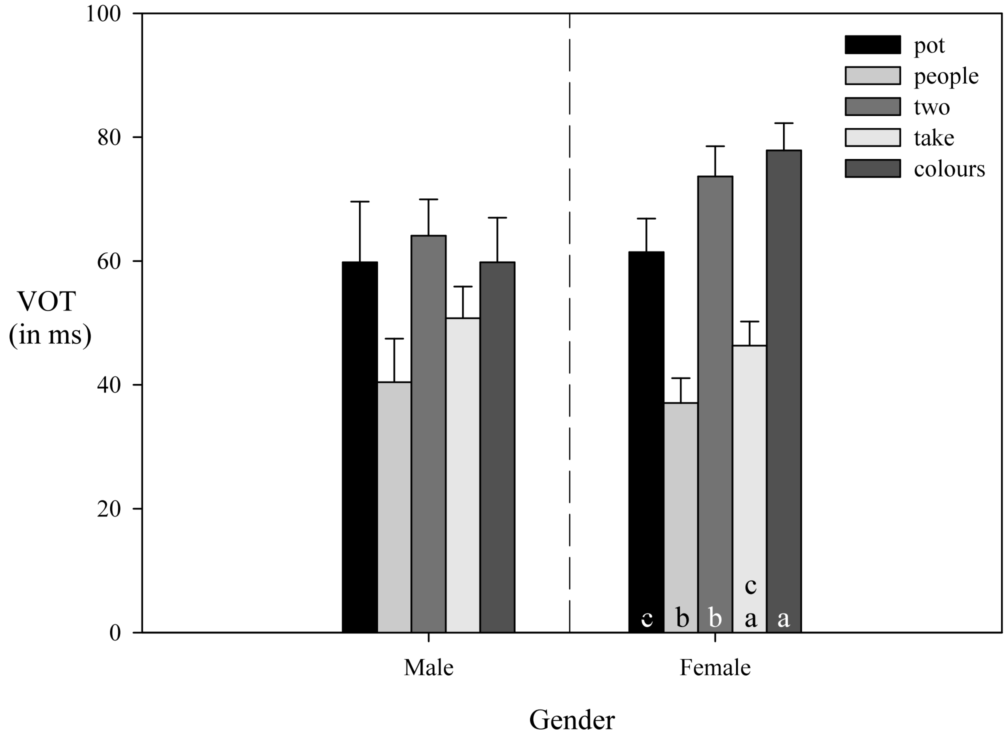


Figure 23. Word effect on VOT measured from selected words.

Means and standard error of means for the VOT measures obtained from selected words ("pot", "people", "two", "take", and "colours") embedded with different consonants (/p/, /t/, /k/) and/or different vowels in the male and female groups separately [male (n = 10), female (n = 13)]. Significantly different pairs in each data set are marked with different letters.

3.2 Perceptual Measures

Results from a series of one-way (H1-H2 level) ANOVAs conducted on the averaged listener responses to the male and female "sentence-embedded vowels" and "sustained vowels" presented in the "vowel identification" and "clarity discrimination" tasks are described in this section. Means and standard deviations of the scores for each of the five H1-H2 levels are illustrated to show whether there is a trend that goes with the expectation that the lowest level (Level 1) of H1-H2 amplitude difference would achieve the lowest percentage of correct vowel identification and percentage of "perceived as clearer" and that the scores would increase with increasing level of H1-H2 amplitude difference.

3.2.1 Vowel Identification Task

Results for the first set of male voice stimuli for the vowels /i/ and /o/ showed the expected trend to a degree although the gains in intelligibility for levels of the vowel /i/ were not statistically significant (see Figure 24). The Levels 4 and 5 of the vowel /o/ received significantly higher percent correct scores than the Level 1 for that vowel although it had an unexpectedly high intelligibility rating. Results for the vowel /a/ failed to show the expected trend for the Level 1 stimuli which received percentage correct rating of over 90 % though the Level 2 stimuli received a significantly poorer rating than stimuli of Levels 3, 4, and 5 which may indicate an intelligibility threshold somewhere between the H1-H2 amplitude difference levels of the Levels 2 and 3 stimuli if the Level 1 stimuli is disregarded. Results for the vowel /u/ showed a percentage correct rating just short of 50% rating for the Level 1 stimuli and ratings beyond 60% for Levels 2 through to 5 (see Figure 24).

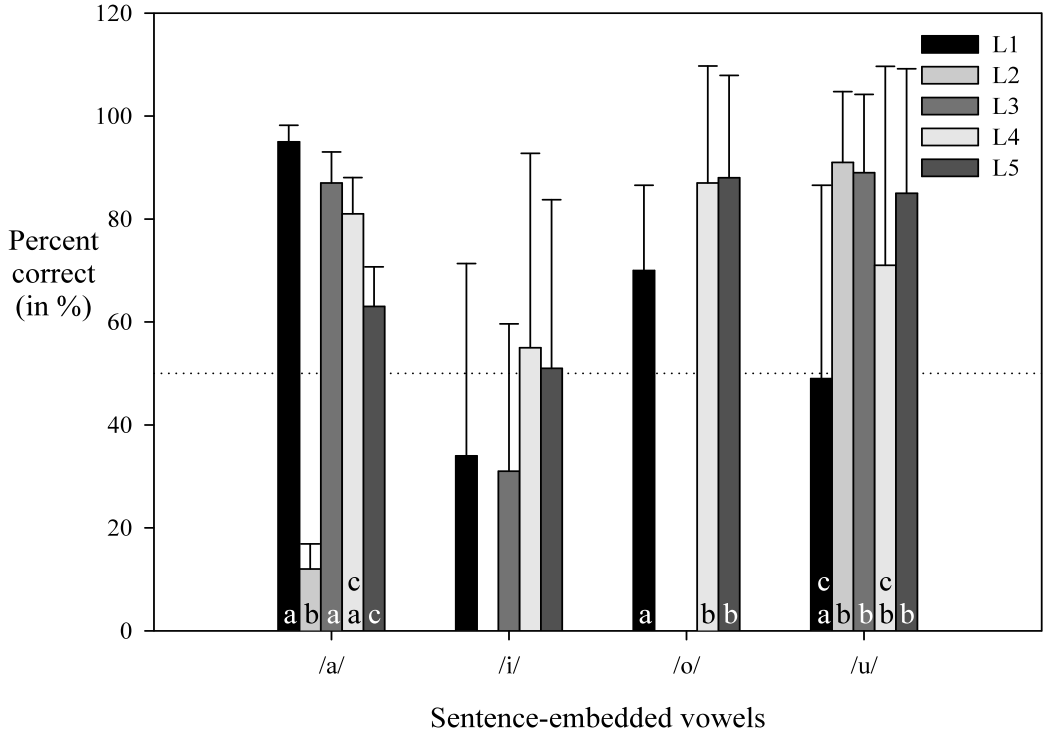


Figure 24. Averaged listener responses to the male "vowel identification" task first set.

Means and standard deviations of percentage correct results (n = 20) calculated from the averaged listener responses to the first set of "male sentence-embedded vowels" used in the "vowel identification" task. Levels L1 to L5 represent increases in H1-H2 amplitude difference with L1 being the lowest level. Significantly different levels in each data set are marked with different letters.

For the second set of male voice stimuli, the vowel /o/ also showed the expected trend of increased intelligibility with decreasing breathiness over the three levels represented with stimuli although the differences in levels were not statistically significant (see Figure 25). The vowel /a/ showed an increase in intelligibility from Level 1 to 2 that was not large enough to be statistically significant (see Figure 25). Results for the vowels /i/ and /u/ failed to show the expected trend (see Figure 25).

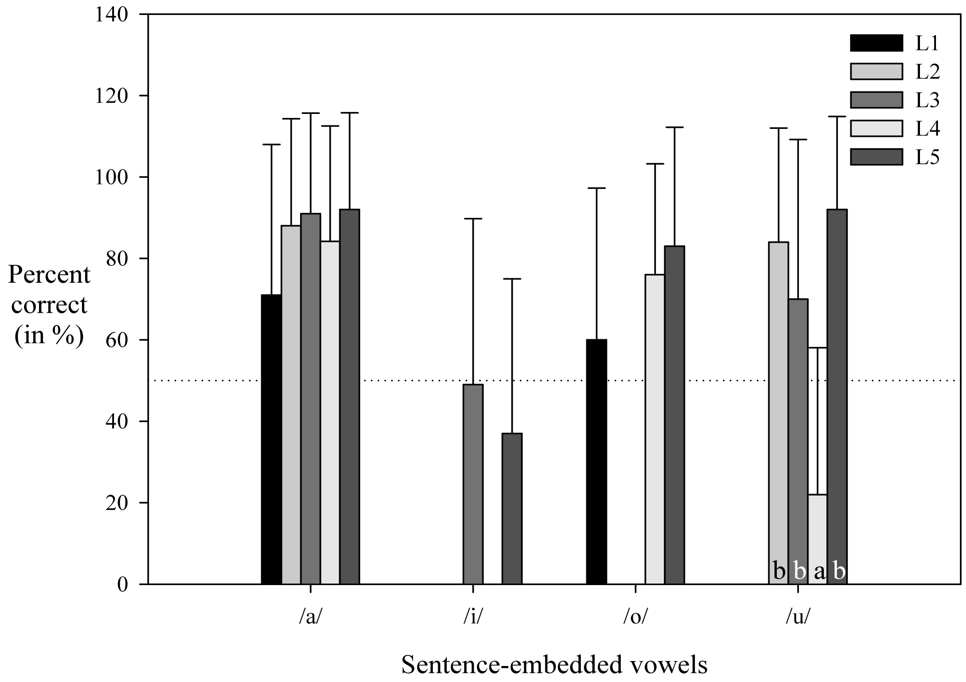


Figure 25. Averaged listener responses to the male "vowel identification" task second set. Means and standard deviations of percentage correct results (n = 20) calculated from averaged listener responses to the second set of "male sentence-embedded vowels" used in the "vowel identification" task. Levels L1 to L5 represent increases in H1-H2 amplitude difference with L1 being the lowest level. Significantly different levels in each data set are marked with different letters.

Results from the first set of female voice "sentence-embedded vowels" failed to show the expected trend. For the first set of female voices, it is shown in Figure 26 that no significant level difference on the percent of correct vowel identification was found for the vowels /i/ and /o/. For the vowel /a/, the percent of correct vowel identification for Level 4 was lower than 50% and significantly lower than all the other levels (see Figure 26). For the vowel /u/, the percent of correct vowel identification for Level 5 was above 50% but significantly lower than that of Levels 1 and 4 (see Figure 26).

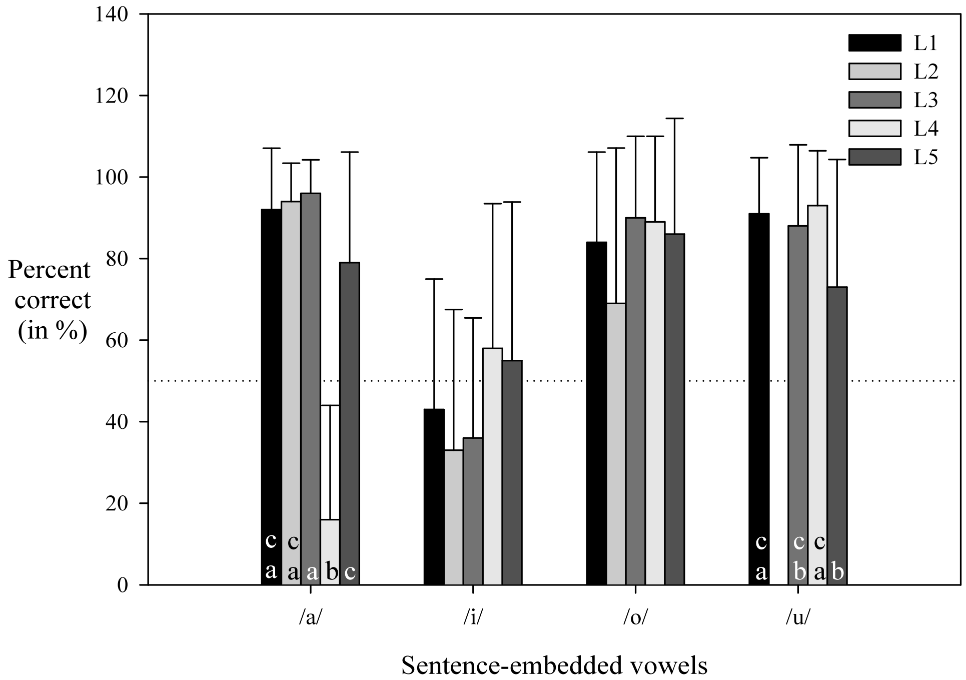


Figure 26. Averaged listener responses to the female "vowel identification" task first set.

Means and standard deviations calculated from averaged listener responses (n = 20) to the first set of "female sentence-embedded vowels" used in the "vowel identification" task. Levels L1 to L5 represent increases in H1-H2 amplitude difference with L1 being the lowest level. Significantly different levels in each data set are marked with different letters.

The second set also failed to show the expected trend (Figure 27). For the second set of female voice "sentence-embedded vowels" used in the "vowel identification" task, the levels were not significantly different for the vowel /a/ (see Figure 27). For the vowel /i/, Level 4 was significantly lower than Levels 3 and 5 (see Figure 27). For the vowel /o/, Level 2 was significantly lower than all the other levels. For the vowel /u/, Level 3 was significantly lower than Levels 4 (see Figure 27).

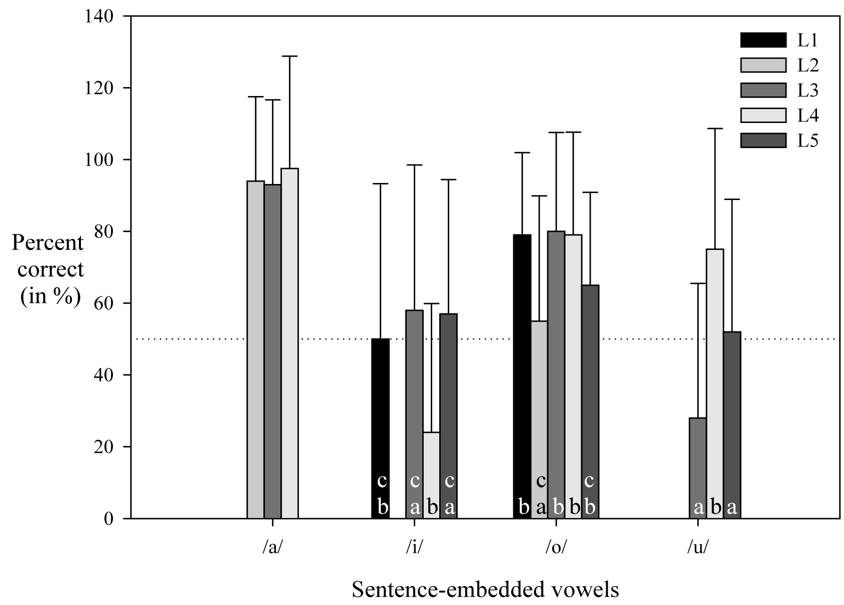


Figure 27. Averaged listener responses to the female "vowel identification" task second set.

Means and standard deviations calculated from averaged listener responses (n = 20) to the second set of "female sentence-embedded vowels" used in the "vowel identification" task. Levels L1 to L5 represent increases in H1-H2 amplitude difference with L1 being the lowest level. Significantly different levels in each data set are marked with different letters.

Results for the male vowel /a/ from the presentation of a small set of "sustained vowels" in the same task did not show the expected trend. The Level 1 which ought to have been the breathiest stimuli in the set received a 100 % correct rating, the Level 2 stimuli a rating of about 70% and further but lesser increase in intelligibility of about 20 percentage points for the Level 4 stimuli (see Figure 28). The vowel /i/ from the same task and stimuli type did not show the expected trend with the breathiest stimuli receiving an averaged percentage correct score of over 70 % (see Figure 28).

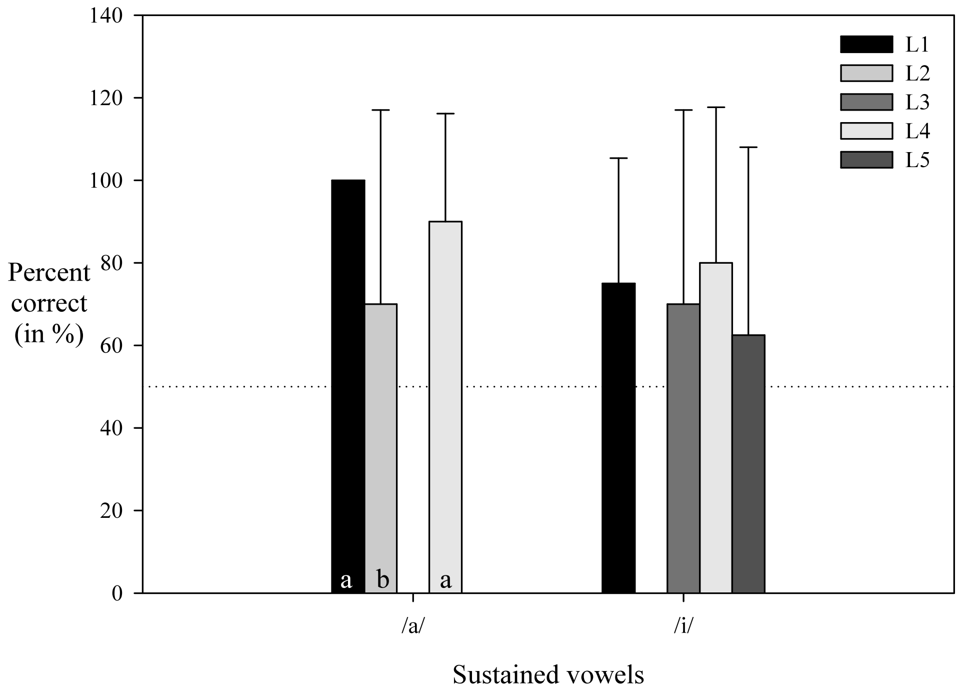


Figure 28. Averaged listener responses to the male "vowel identification" with "sustained vowels".

Means and standard deviations calculated from averaged listener responses (n = 20) to a small set of "male sustained vowels" used in the "vowel identification" task. Levels L1 to L5 represent increases in H1-H2 amplitude difference with L1 being the lowest level. Significantly different levels in each data set are marked with different letters.

The small sample of female voice “sustained vowels” used for the same “vowel identification” task showed the expected trend for the vowel /a/ (see Figure 29). For the vowel /a/, the first level of stimuli was rated as fairly intelligible with a rating of 70%, yet it was the poorest rating stimuli, more than 20 percentage points below Levels 2, 3, 4 and 5. The levels for the vowel /i/ did not show the expected trend with Level 1 having the highest intelligibility rating and Level 5 the lowest intelligibility rating. There was no significance between differences in Levels 1, 2 and 5 for the vowel /i/ (see Figure 29).

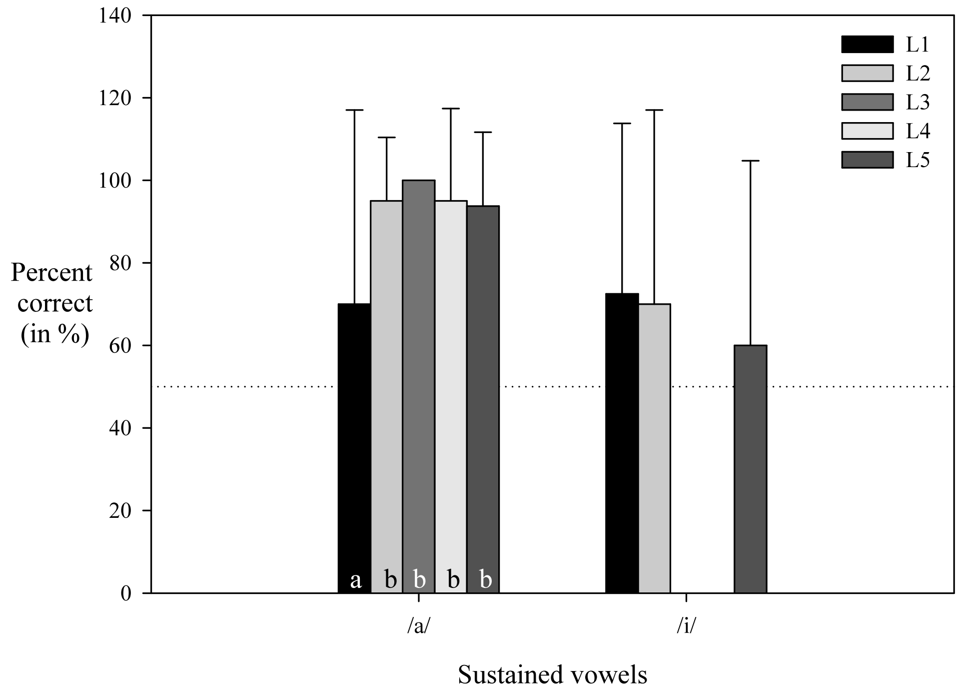


Figure 29. Averaged listener responses to the female “vowel identification” task with “sustained vowels”.

Means and standard deviations calculated from averaged listener responses (n = 20) to a small set of “female sustained vowels” used in the “vowel identification” task. Levels L1 to L5 represent increases in H1-H2 amplitude difference with L1 being the lowest level. Significantly different levels in each data set are marked with different letters.

3.2.2 Clarity Discrimination Task

Means and standard deviations calculated to determine the effect of the level of the H1-H2 amplitude difference measure used averaged listener responses to “sustained vowels” from the “clarity discrimination” task.

Results for the first set of male voice stimuli conformed to the expected trend of increasing clarity for the first three of four levels for the vowel /a/. There was a steep increase in clarity of the magnitude of 40% between stimuli of Levels 2 and 3 (see Figure 30). Results for the vowel /i/ did not showed the expected trend with a decline in clarity evidenced for increasing levels of H1-H2 amplitude difference although stimuli for Levels 1 and 2 in which breathiness is most likely to be evidenced were not represented (see Figure 30).

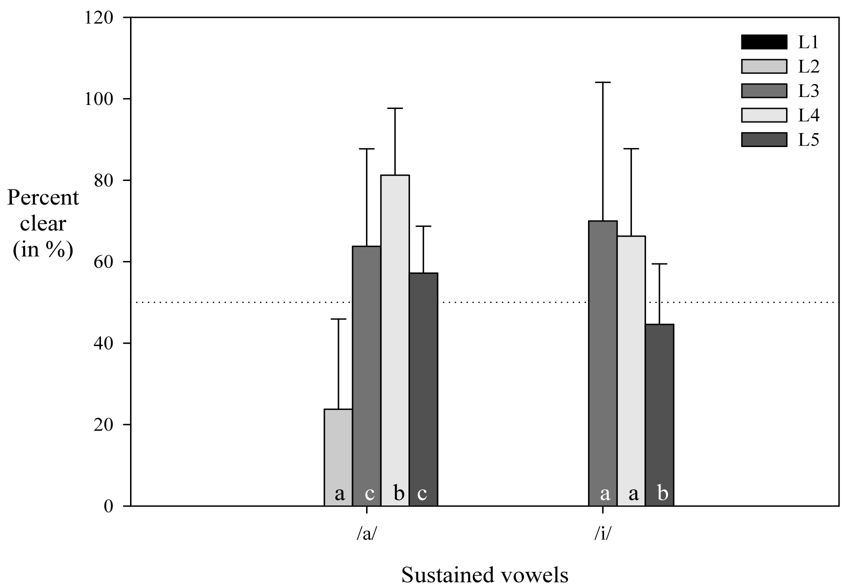


Figure 30. Averaged listener responses to the male “clarity discrimination” task first set.
Means and standard deviations calculated from averaged listener responses (n = 20) to the first set of “male sustained vowels” used in the “clarity discrimination” task. Levels L1 to L5 represent increases in H1-H2 amplitude difference with L1 being the lowest level. Significantly different levels in each data set are marked with different letters.

Results for the vowel /a/ from the second set of male "sustained vowel" stimuli also showed a steep gain clarity of over 60 percentage points occurred at a point between the H1-H2 amplitude difference levels for Level 2 and 3 stimuli, although the Level 1 stimuli with its 100 % clear is most unexpected (see Figure 31). Results for the vowel /i/, limited to two stimuli, did not show the expected trend, with a reduction in clarity occurring between Levels 3 and 4 of more than 25 percentage points (see Figure 31). For the vowel /a/ of the second set of "male sustained vowels", Level 2 was significantly different in comparison with Levels 4 and 5, as was Level 3 in comparison with Levels 2, 4 and 5 (see Figure 31). Level 4 was not significantly different to Level 5 (see Figure 31). For vowel /i/ of the second set of "male sustained vowels", Level 3 was significantly different in comparison with Level 4 (see Figure 31).

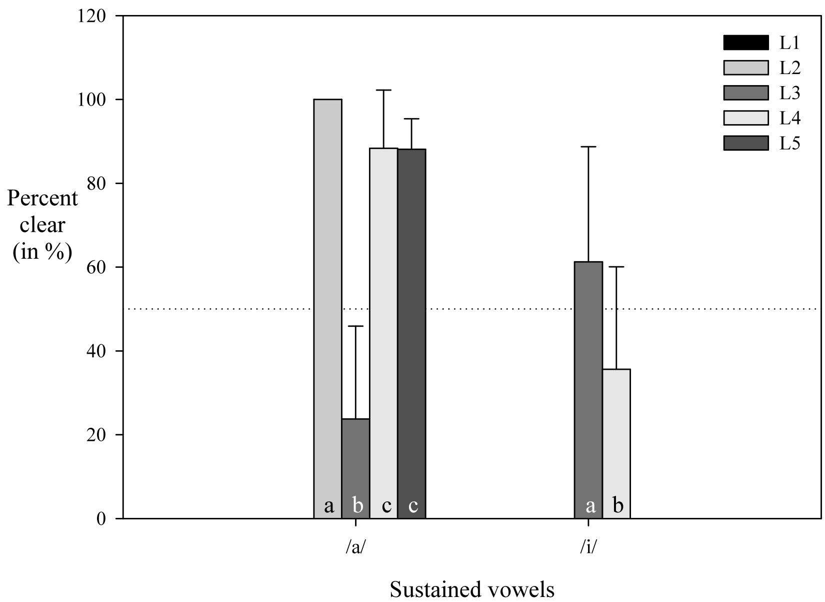


Figure 31. Averaged listener responses to the male "clarity discrimination" task second set.
Means and standard deviations calculated from averaged listener responses (n = 20) to the second set of "male sustained vowels" used in the "clarity discrimination" task. Levels L1 to L5 represent increases in H1-H2 amplitude difference with L1 being the lowest level. Significantly different levels in each data set are marked with different letters.

The small sample of "male sentence-embedded vowels" presented in the "clarity discrimination" task showed increasing clarity for the first three levels of the vowel /a/ (see Figure 32). The clarity rating increased from Level 1 which achieved just over 50% clarity to a score of 95% clarity at Level 3, followed by a decline in clarity of just short of 20 percentage points at Level 4 (see Figure 32). Results for the vowel /i/ conformed to the expected trend with increasing clarity shown for the three levels represented although there was no significant level effect.

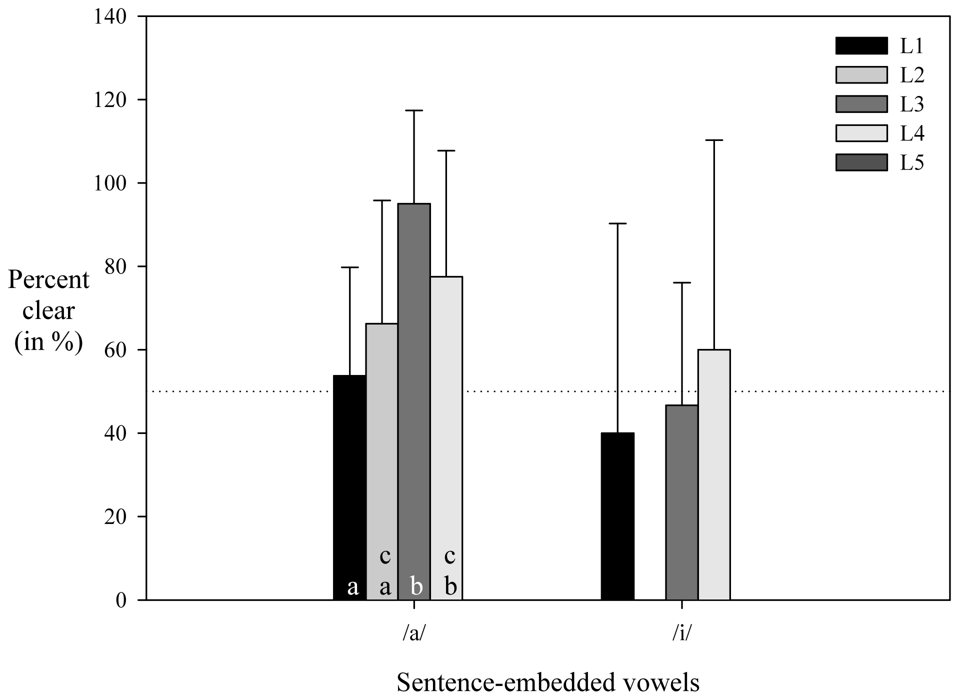


Figure 32. Averaged listener responses to the male "clarity discrimination" task with "sentence-embedded vowels".

Means and standard deviations calculated from averaged listener responses (n = 20) to a small set of "male sentence-embedded vowels" used in the "clarity discrimination" task. Levels L1 to L5 represent increases in H1-H2 amplitude difference with L1 being the lowest level. Significantly different levels in each data set are marked with different letters.

Results for the first set of “female sustained vowels” /a/ and /i/ from the “clarity discrimination” task failed to show the expected trend (see Figure 33). Level 2, the first level represented by stimuli for the vowel /a/ proved to be the most clear with a rating of 90% clarity whereas Level 4 received a clarity rating below 50% (see Figure 33). For vowel /a/ clarity ratings, Level 2 was significantly higher from Levels 3, 4, and 5 while Level 4 was significantly lower than Levels 3 and 5, which were not significantly different. For vowel /i/, Level 3 was significantly higher than Levels 2 and 5, which were not significantly different (see Figure 33).

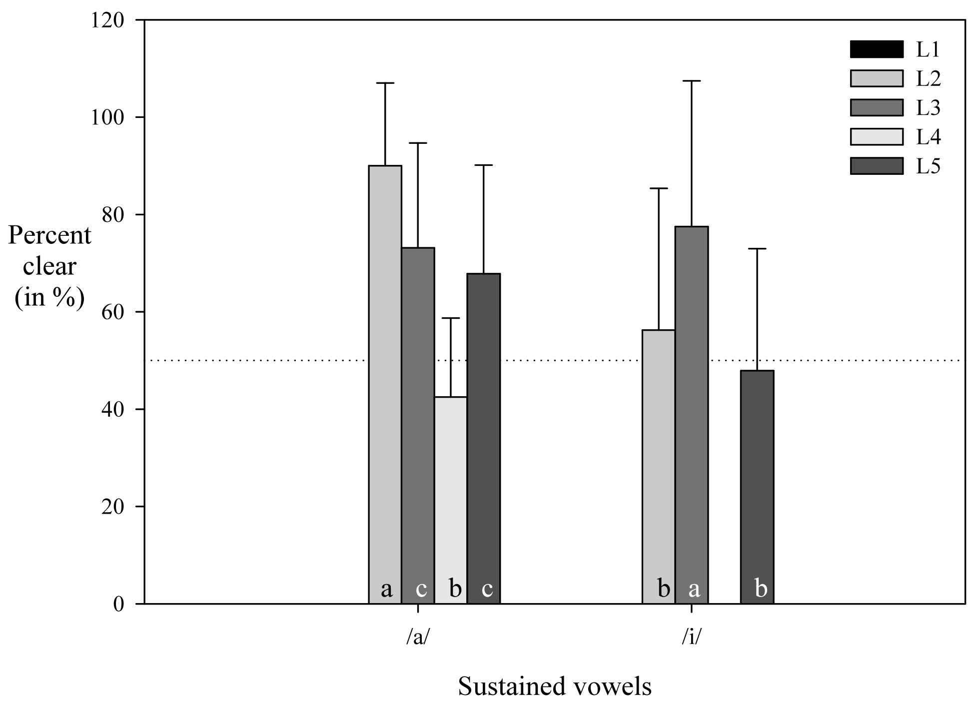


Figure 33. Averaged listener responses to the female “clarity discrimination” task first set. Means and standard deviations calculated from averaged listener responses (n = 20) to the first set of “female sustained vowels” used in the “clarity discrimination” task. Levels L1 to L5 represent increases in H1-H2 amplitude difference with L1 being the lowest level. Significantly different levels in each data set are marked with different letters.

Results for the second set of "female sustained vowels" for the "clarity discrimination" task also failed to show the expected trend (see Figure 34). For the vowel /a/, Level 5 of the sustained /a/ stimuli achieved a mere 50% clarity rating although there was no significant difference between Levels 4 and 5. For the vowel /i/, Level 3 dropped 20 percentage points in clarity from that of the Level 2 (see Figure 34).

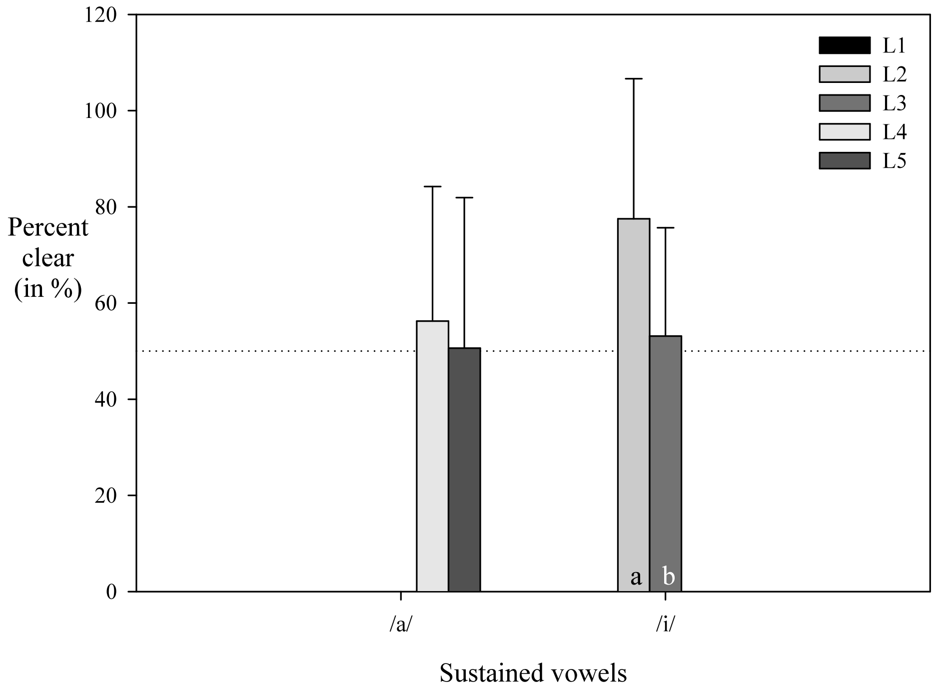


Figure 34. Averaged listener responses to the female "clarity discrimination" task second set.

Means and standard deviations calculated from averaged listener responses (n = 20) to the second set of "female sustained vowels" used in the "clarity discrimination" task. Levels L1 to L5 represent increases in H1-H2 amplitude difference with L1 being the lowest level. Significantly different levels in each data set are marked with different letters.

The small sample of “female sentence-embedded vowels” presented in the “clarity discrimination” task showed increasing clarity for Levels 1 to 3 of the four levels represented and then an unexpectedly steep drop in clarity of 50 percentage points for Level 4 for the vowel /a/ (see Figure 35).

Results for the vowel /i/ for Levels 1, 2, and 3 all received low clarity ratings with Level 2 receiving the highest rating of just short of 30% clarity from which point the clarity increases steeply to a 55% clarity rating received by the Level 4 stimuli (see Figure 35). For the vowel /a/, the Level 3 stimuli were significantly clearer than Levels 1 and 4 (see Figure 35). Level 1 did not receive a significantly different clarity rating from Levels 2 and 4 nor did Level 2 receive a rating of significantly improved clarity from the ratings of Levels 3 and 4 (see Figure 35). For the vowel /i/, the Level 4 stimuli showed statistical significance in the degree to which it was clearer than Levels 1 and 3, though not in the extent to which it was clearer than the Level 2 stimuli. The Level 1 stimuli was not significantly different in clarity from stimuli of Levels 2 and 3, whilst Level 2 did not have significantly different degrees of clarity than Levels 3 and 4 (see Figure 35).

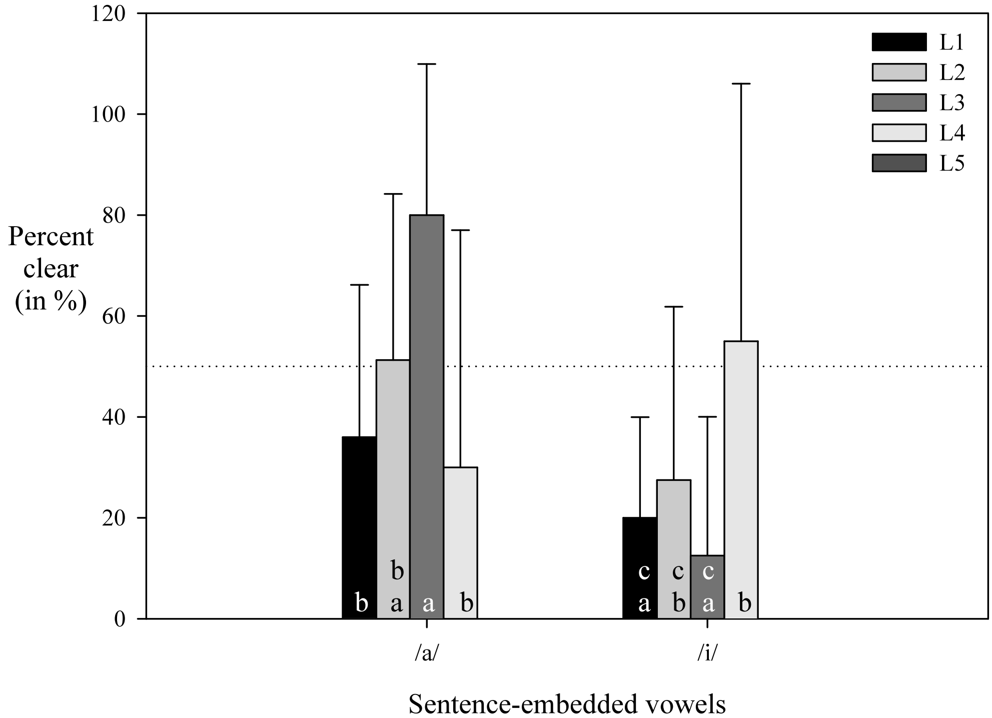


Figure 35. Averaged listener responses to the female "clarity discrimination" task with "sentence-embedded vowels".

Means and standard deviations calculated from averaged listener responses (n = 20) to a small set of "female sentence-embedded vowels" used in the "clarity discrimination" task. Levels L1 to L5 represent increases in H1-H2 amplitude difference with L1 being the lowest level. Significantly different levels in each data set are marked with different letters.

3.3 Summary of Main Findings

The main findings of the acoustic study are:

1. **Formant frequencies and vowel space:** There was a glottal closure effect for the F2 frequency measured from "male sustained vowels". The F2 frequency averaged across vowels /a/ and /i/ was significantly higher for incomplete glottal closure than complete glottal closure. Glottal closure by vowel interaction effect showed that only for the condition of complete glottal closure the F2 frequency was significantly higher for the vowel /i/ than for the vowel /a/.
2. **H1-H2 amplitude difference:** There was a glottal closure effect for "male sentence-embedded vowels". The H1-H2 amplitude difference measure was significantly lower (i.e., greater H1 dominance) for the condition of incomplete glottal closure than for the complete glottal closure condition. As it is the lower H1-H2 amplitude difference measures that are found consistent with a dominant H1 and breathiness in the literature, the finding of a lower H1-H2 in the group with incomplete glottal closure in comparison with the group with complete glottal closure supports, although only with male data, the hypothesis that the H1-H2 measure would be useful for differentiating pathological voice with and without complete glottal closure.
3. **Singing power ratio:** Whilst there was no significant difference between glottal closure conditions for the SPR measures, there was a significant effect of interaction between glottal closure and vowel for the "male sustained vowels". For males, the sustained vowel /i/ had a significantly higher SPR measure (i.e., stronger voice projection power) than sustained /a/ within the incomplete glottal closure condition.

4. **Consonant-to-vowel energy ratio and VOT:** There was an absence of glottal closure effect for the CV energy ratio and VOT measures taken from "sentence-embedded words".

The main findings of the perceptual study are:

1. **Vowel identification:** Percentage correct results from the vowel identification task for some vowels, /a/ and /o/ in male voice in particular, indicated the presence of a threshold below which breathiness impacts harshly on intelligibility and above which breathiness has little effect on intelligibility. This is evidenced by the contrast between the large gains in intelligibility achieved in percentage correct ratings of vowel stimuli of the lower levels of H1-H2 amplitude difference and the generally higher intelligibility ratings achieved by the stimuli with higher levels of H1-H2 amplitude difference. As the levels of H1-H2 amplitude difference rose above a certain point less gain in intelligibility was made.
2. **Vowel clarity:** Percentage clear results from the "clarity discrimination" task yielded results that conformed to the expected pattern of increased clarity with decreased breathiness only for some levels of some vowel stimuli. An increased clarity with increased H1-H2 level is observed for both genders in sentence-embedded vowels. For sustained vowels, this trend is only evident for the vowel /a/ in male voice.

4 Discussion

The purpose of this study was to add to the body of literature investigating the relationship between voice characteristics and speech intelligibility. This section provides a discussion of the findings of this study in relation to the research questions and previous research. Clinical implications, limitations, and directions for future research are also discussed.

4.1 Study Findings in Relation to the Hypotheses

The overall hypothesis of this study was that the acoustic measures F1 and F2 frequencies, H1-H2, SPR, CV energy ratio, and VOT would differentiate between samples of pathological voice produced by voice patients with and without complete glottal closure, and that perceptual measures would confirm the relationship between the breathiness-related measure with the perception of vowel identification and clarity. The general finding from the acoustic study is that voice produced by voice patients with incomplete glottal closure, in comparison with voice produced by voice patients with complete glottal closure, is associated with an increase in F2 as shown in male sustained vowels, decreased H1-H2 (i.e., greater H1 dominance) as shown in male sentence-embedded vowels, and greater vowel difference in SPR (with /i/ showing greater SPR than /a/ in male sustained phonation). The relationship between H1-H2 level and the perception of vowel identification is not linear although there is some indication that lower H1-H2 level (i.e., greater H1 dominance) is associated with reduced intelligibility and clarity scores.

The finding of a significant glottal closure or glottal closure by vowel interaction effect found in F2, H1-H2, and SPR supports the hypothesis that these measures would be sensitive in detecting a voice difference related to breathy voice. The hypothesis is only partly supported, however, because only male voice yielded a significant finding, suggesting that the complete and incomplete glottal closure conditions in female voice may be harder to differentiate based on the acoustic features related to breathiness. As mentioned in the literature review, the female larynx is predisposed with a laryngeal structure more likely to produce breathy voice. As breathy voice or incomplete glottal closure are common in females, the distinction between pathological voice with and without complete glottal closure may be easily masked by the greater normal variations in the degree of glottal closure.

The hypothesis that voice quality may have an impact on speech intelligibility and thus breathy voice may be associated with changes in F1 and F2 frequencies resulting in poorer vowel differentiation is partly supported by the finding of a loss of difference between /a/ and /i/ on F2 frequency in the sustained phonation of males with incomplete glottal closure.

The hypothesis that voice associated with incomplete glottal closure would show a more dominant H1 is supported by the finding in male sentence-embedded vowels. The hypothesis that SPR would be useful for detecting the voice difference related to glottal competence is only partly supported by the finding in the male sustained phonation that a greater SPR difference between /a/ and /i/ was found in the incomplete glottal closure condition. The hypotheses about the usefulness of CV energy ratio and VOT in differentiating voice associated with complete and incomplete glottal closure are not supported. These findings suggest that pathological voice with and without complete glottal closure may share some common signs of voice changes and thus they are not readily differentiable based on this selection of acoustic measures.

As for the perceptual study, it was expected that a dominant H1 associated with breathiness and thus lower H1-H2 level would result in the poorest percentage correct scores. This would result in bar graphs for both the "vowel identification" task and "clarity discrimination" task that would have increasing bar height with increasing H1-H2 levels.

This pattern can be seen in isolated results such as for the vowel /o/ of the first and second set of “male sentence-embedded vowels” in the “vowel identification” task (see Figure 24 and Figure 25) but the pattern when shown is not linear. There are considerably larger gains in intelligibility made within stimuli at the lower H1-H2 levels than within stimuli at the higher H1-H2 levels. This suggests that extreme levels of breathiness may severely curtail speech intelligibility whereas smaller amounts of breathiness have considerably less impact on speech intelligibility.

4.2 Study Findings in Relation to Previous Research

The finding that the F2 frequencies of male sustained phonation were differentiated between /a/ and /i/ in the complete glottal closure condition but not in the incomplete glottal closure condition provides evidence that breathiness may adversely impact on speech intelligibility. As reduced vowel differentiation may be represented by a more compressed vowel space, the present study may be related to previous studies of the relationship between vowel space and speech intelligibility. For example, Bradlow et al. (1996) studied normal voice and found that measures of vowel space area and dispersion calculated from the F1 and F2 frequencies were highly correlated with intelligibility scores calculated from sentence transcriptions produced by the listeners (Bradlow et al., 1996). A further measure of the range in F1 and F2 frequencies (the difference between maximum and minimum values of each) was calculated to determine which formant frequency could be credited with the high correlation. It was found that F1 was the better correlate of intelligibility which was explained by English vowels having greater distinctions relating to height (of which F1 is the more important acoustic correlate) than front-backness (of which F2 is the more important acoustic correlate) (Bradlow et al., 1996). In contrast, a study of the speech intelligibility of adolescents with severe-to-profound hearing impairment found F2 to be highly correlated with intelligibility (Monsen, 1978). Specifically, it was found that the increase in the range of F2 frequencies measured from the vowels /i/ and /ɔ/ did cause an expansion of vowel space correlating with increased intelligibility ratings (Monsen, 1978). As formant frequencies shown to be related to speech intelligibility in the literature are found in this study to be susceptible to the pathological voice associated with incomplete glottal closure, the present finding adds to the literature showing the relationship between vocal control and speech intelligibility.

The finding that the H1-H2 amplitude difference measure was significantly smaller (i.e., more breathy) for the condition of incomplete glottal closure than for complete glottal closure is consistent with the common finding from Bickley (1982), Fischer-Jorgensen (1967), Hillenbrand et al. (1994), Klatt and Klatt (1990), and Ladefoged (1983) that a dominant first harmonic is associated with breathy vowels. Findings from the perceptual judgments investigated in the second part of the study confirm that the greater levels of H1-H2 amplitude difference (represented by smaller values and more dominant first harmonics) did impact negatively on the intelligibility and clarity of the vowels.

According to Klatt and Klatt (1990), females on average have a greater tendency than males to exhibit breathy voices in the normal speaking population either due to physiological or functional reasons. The present finding shows that a statistical significant difference on H1-H2 amplitude difference between the complete and incomplete glottal closure groups was found in the male voice samples only. As breathiness is less common in normal male voices, it is likely that the difference between voices associated with complete and those with incomplete glottal closure is more evident in the male voice.

Hanson and Emanuel (1979) referred to selecting "non high" vowels for the measurement of H1-H2 amplitude difference so that H1 and F1 were well separated.

Likewise Henton and Bladon (1985) selected only open vowels from which to measure H1-H2 amplitude difference to avoid interference from the first formant. The explanation was that only open vowels had first formants high enough not to increase the amplitudes of the lower harmonics (Henton & Bladon, 1985). Open vowels are the equivalent of low vowels. Vowel height is inversely related to the frequency of F1, the higher the tongue position the lower the F1 frequency (Reetz & Jongman, 2009). Of the vowels used in this study /i/ and /u/ would be considered high vowels while /a/ and /o/ are low vowels. It is possible that H1-H2 amplitude difference is most usefully measured from low vowels. It can be seen from the results of the "clarity discrimination" task that the increase of clarity scores as a function of H1-H2 level is more clearly shown for the vowel /a/, suggesting that the effect of F1 on H1 may affect the impact of breathiness on the perception of vowel clarity.

The finding of a significant glottal closure by vowel interaction effect from "male sustained vowels" but not from "sustained vowels" may be related to the task difference. Task effect on the SPR measure has been reported in the literature. For example, Omori (1996) found that SPR measured from sung /a/ samples was significantly greater than that measured from spoken /a/ samples produced by trained singers (whose vowel samples had significantly stronger SPR measures than non singers). The present finding suggests that sustained phonation may be more sensitive in detecting voice changes related to SPR.

Findings from the "clarity discrimination" task failed to reveal a consistent pattern about the relationship between H1-H2 level and the perception of vowel clarity. Studies aiming to correlate acoustic features with perceptual qualities or pathologies have often yielded results that are ambiguous sometimes contradictory to the extent that the selection of appropriate measures and their interpretation remain unclear (Michaelis, Fröhlich, & Strube, 1998). One contributing factor is likely to be that most acoustic measures are sensitive to more than one acoustic property. For example, jitter and shimmer typically used to detect aperiodicity are also sensitive to additive noise in speech signals (Michaelis et al., 1998). The challenge is to find the measures that provide the best independent assessment of different acoustic features (Michaelis et al., 1998).

4.3 Clinical Implications

This study shows that a large enough change in H1-H2 amplitude difference, which ought to result in increased breathiness, has a detrimental effect on the intelligibility of speech. The incomplete closure of the glottis that occurs with breathy voicing ensures continuous airflow throughout the glottal cycle which contrasts with the glottal cycle of a normal voice that achieves complete closure during periodic closure phases (Baart, 2010). Aspiration noise is produced by the additional airflow when combined with voicing (Titze, 1994). It is present especially in the higher frequencies of the spectrum of vowels (Klatt & Klatt, 1990). The contribution of aspiration noise to speech can be likened to the effect of background noise on speech perception such that intelligibility may suffer with decreases in the signal-to-noise ratio (Henton & Bladon, 1985). Knowledge gained about the impact of breathiness on the intelligibility of speech could be used to improve speech synthesis and speech recognition algorithms.

A consequence of sensorineural hearing loss (SNHL) is the reduced ability to hear in background noise (Arlinger, 2003). Damage to cochlear outer hair cells reduces the capacity to resolve frequency information making it difficult to distinguish speech signals from background noise (Venema, 2006). The more severe the hearing loss, the greater are the requirements for favourable SNRs (Crandell & Smaldino, 2002). A person with SNHL is likely to have greater difficulty understanding breathy speech than a person with normal hearing. Aspiration noise would act to mask the speech signal. This would disadvantage a person with SNHL who had a frequent conversational partner with a particularly breathy voice. The current hearing aid technology solutions that effectively increase signal-to-noise ratio, including directional microphones and FM systems, rely on the noise and the signal having different sources. They would not improve the perception of a breathy voice in which the signal and noise share a common source. A better understanding of the effect of voice quality on speech intelligibility would help develop an effective hearing aid sound processing scheme that alters gain for selected portions of the speech signal to optimise a hearing-impaired listener's ability in speech recognition (Smith & Levitt, 1999).

4.4 Limitations and Future Directions

Picheny et al. (1986) considered that the usefulness of CV energy ratio data is obscured when measured from sentence material due to the variable vowel intensity in contrast to more stable vowel intensity in isolated words or syllables. Using nonsense sentences as speech material, they measured the intensities of consonants in isolation rather than CV energy ratios considering this to be a more valid measure (Picheny et al., 1986).

Monsen (1978) in his study of acoustic measures and intelligibility scores of hearing impaired hearing adolescents, used list words rather than sentence materials to measure VOT in order to maintain stable phonological and prosodic context. In contrast, this study used words from sentence readings. It is feasible that list words would have been more appropriate for the same reasons yet it is also the case that with normal hearing talkers, speech material may have been subject to less variability than for talkers with hearing impairment.

As the effect of voice quality on vowel space would be more meaningfully compared within an individual than between individuals, the vowel space changes as a function of the level H1-H2 amplitude difference may be hard to interpret due to a lack of control of individual differences. The vowel space plots and areas pertaining to this study, displaying changes as a function of the level H1-H2 amplitude difference, are shown in Appendices 8 to 14. Within-subject comparisons should be conducted in the future to clarify the impact of breathiness on vowel space area.

A limitation of this study was brought about by the manner by which stimuli were selected for the perceptual study. Having four groups of "sentence-embedded" stimuli and two groups of the "sustained vowel" stimuli, each with different ranges of levels of H1-H2 amplitude difference, made interpretation of and comparison between stimuli in the perceptual study problematic. An uneven spread of values within the ranges meant that selection for some of the levels was lacking or limited to one sample. The use of equal intervals to determine the ranges from which the selection of stimuli was made was the major limiting factor here. A method of selection that accounted for the uneven spread of values may have yielded results that are more useful.

Further limitations of the study may be revealed by the selection of choices made by the listeners. The averaged selection choices made by the listeners in response to the request that they select the vowel closest to the vowel that they heard in the "vowel identification" task when the stimuli presented was /i/ are shown in Appendices 15 and 16. It is shown in Appendices 15 (for responses to the male /i/) and 16 (for female /i/) that the vowel /i/ was often mistaken for /e/. There were four vowels used as stimuli for the correct identification task, namely /a, i, o, u/. The listeners were given five options with the addition of the vowel /e/ from which to select the vowel that most closely resembled the vowel they heard. The additional vowel was added as a reliability measure. If a listener always favoured the one vowel not represented among the stimuli then the results from that listener should be disregarded. However, the listeners appeared to confuse /i/ for /e/ (see Appendices 15 and 16). In this instance of confusion, the addition of the vowel /e/ as an option for selection or the use of inexperienced listeners as participants for the study may have contributed to the low intelligibility scores for the vowel /i/. Inexperienced listeners are less likely to be familiar with linguistic symbols and therefore less likely to make the correct distinction between letters and phonemes so when presented with the sound /i/ they may have selected /e/ from the screen in front of them, mistaking it for the letter “e” that sounds like the phoneme /i/. Had the listeners not had the additional vowel to select from or had all participants had a background in speech pathology or linguistics, the results for the vowel /i/ in the “vowel identification” task may have been different.

The layout of the interface screen used for the perceptual tasks may have contributed to the confusion between /i/ and /e/ for as can be seen in Appendix 6, /i/ and /e/ were positioned next to each other. One of the study participants mentioned another possible limitation with regards to the interface screen layout. He thought that the right sided placement of the start button, instead of a middle position, had the potential to create a selection bias in favour of vowels choices toward the right of the screen. Perhaps he was correct. There was a greater distance for the mouse to travel from the start button to the /i/ than to the /e/ (see Appendix 6).

Further investigation into participant vowel selection for the first set of the “vowel identification” task showed that participants most often mistook the “male sentence-embedded vowel” stimuli /a/ for /o/. Participants made the correct selection of the vowel /a/ 85% of the time and mistook it for the vowel /o/ 12% of the time (see Appendix 17). The experimenter listened to the 50 ms sectioned stimuli sample and found that the /a/ did sound like an /o/. In addition, the word “arch” in the sentence recording from which it was sectioned sounded like the word “arch” to the researcher. The researcher, an audiology masters student with no background in Speech Pathology, found that the patient’s voice did not sound normal. Examination at the time of the voice recording revealed that the patient had incomplete glottal closure and had not at that stage received treatment for his voice condition. It is unclear whether this finding is related to the effect of breathiness on vowel intelligibility or an individual's speech error. This reveals a couple of sources of possible study limitations concerning vowel stimuli durations, sample size, and the lack of perceptual judgment as to the presence of breathiness or other qualities of voice in the recordings. Perceptual judgments of voice quality were beyond the scope of this study but might be included in future studies.

Vowel duration was another factor that could have influenced listener judgements. It may have been that the relatively short 50 ms duration of the segments hindered correct identification. It may have been that the use of the mid section, where in theory at least vowels sounds should be their most steady, in preference to sectioning that included phoneme transitional information, hindered correct identification. A downside of the use of longer vowel segments and inclusion of phoneme transitional cues is the potential to cause a ceiling effect by which even the most breathy of stimuli would be perceived as 100% intelligible, thereby masking any useful finding in terms of intelligibility. The reasoning behind the inclusion of the 500 ms sustained vowels in the study was concern expressed by the researcher's audiology colleagues and the researcher herself that brief 50 ms durations might render the vowels unintelligible. However, it can be seen in Figures 24 and 26 from the listener responses to the first set of stimuli along with Figures 25 and 27 from the second set of listener responses that this concern proved groundless for the majority of stimuli.

The researcher was also concerned that many of the 500 ms duration segments sectioned from the sustained vowel recordings sounded as though they were produced with a higher pitch than they would be had they been produced by speech. The use of a small set of "sustained vowels" in the "vowel identification" task (see Figures 28 and 29) and "sentence-embedded vowels" in the "clarity discrimination" task (see Figures 32 and 35) provided a check on their compatibility as stimuli as did the Pearson Product Moment correlations conducted on them in which they were shown to correlate moderately well, as is discussed in further detail in Section 2.2.2 ("Stimuli") and can be seen in Figure 1.

However, the decreasing rates of confusion with increases in levels of H1-H2 amplitude difference, especially marked in the averaged response to the female vowel /i/, do support the main finding of the perceptual study that the listeners percentage of correct responses improved showing greater vowel intelligibility with increasing levels of H1-H2 amplitude difference, levels which ought to result in improved voice quality with less breathiness.

It may be of value to include cepstral measures (mentioned previously in Section 1.2.3.1 ["Acoustic Measures Related to Breathiness in Pathological Voice"]) in future studies relating measures of acoustic parameters from dysphonic speech with intelligibility ratings. Cepstral peak prominence was measured from both sustained vowels and running speech of voice patients in a study by Heman-Ackah et al. (2003) and found to be a good predictor of dysphonia. Cepstral peak prominence and high pass and band pass versions of the same measure proved strong predictors of dysphonia, in particular breathiness and roughness, in a study by Wolfe, Martin, and Palmer (2000) which used multidimensional scaling for the listening tasks. In a meta analysis of individual acoustic measures of overall voice quality by Maryn, Roy, DeBodt, Van Cauwenberge, and Corthals (2009) found that cepstral measures achieved the largest effect size for running speech and among the largest effect sizes for sustained vowels. Cepstral peak measures were found to provide the highest correlations with breathiness ratings from the range of acoustic measures employed in a study by Hillenbrand et al. (1994), even though a fully automatic method of measurement was used which proved to have inbuilt errors of accuracy in comparison with hand measures of the same stimuli. High levels of periodicity should result in a more prominent cepstral peak than less periodic signals, though it is the prominence of the peak that is important for the measure not the absolute amplitude (Hillenbrand et al., 1994). Because the amplitude of the peak is affected by overall energy and window size in addition to periodicity, overall amplitude can be normalised by a number of methods (Hillenbrand et al., 1994).

4.5 Conclusion

In this study, consideration has been given to determine to what extent different levels of breathiness, as evidenced by different levels of the H1-H2 measure, might impact negatively on the intelligibility of the speech of people with various types of voice pathology. The acoustic measures which proved to be sensitive in differentiating between pathological voice with and without complete glottal were the frequency of F2, amplitude difference of the first two harmonics (H1-H2), and SPR. In the second part of the study, vowel samples measured as having different levels of the H1-H2 amplitude difference were presented to listeners to determine to what extent they could be identified and to what degree they were perceived as clear. A trend of showing increasing intelligibility and clarity scores with an increase in H1-H2 (indicating less breathy) has been observed in male voice and mainly in the vowel /a/. There were indications of the existence of thresholds of breathiness beyond which lesser degrees of breathiness had little effect on intelligibility or clarity.

References

Allen, J. S., Miller, J. L., & DeSteno, D. (2003). Individual talker differences in voice-onset-time. *Journal of the Acoustical Society of America, 113*(1), 544-552.

Arlinger, S. (2003). Negative consequences of uncorrected hearing loss - A review. *International Journal of Audiology, 42*(SUPPL. 2), 2S17-12S20.

Auzou, P., Özsancak, C., Morris, R. J., Mary, J., Eustache, F., & Hannequin, D. (2000). Voice onset time in aphasia, apraxia of speech and dysarthria: A review. *Clinical Linguistics and Phonetics, 14*(2), 131-150.

Baart, J. L. G. (2010). *A field manual of acoustic phonetics*. Dallas, Texas: SIL International.

Baum, S. R., & Ryan, L. (1993). Rate of speech effects in aphasia: Voice onset time. *Brain and Language, 44*(4), 431-445.

Bickley, C. (1982). Acoustic analysis and perception of breathy vowels. *Speech Communication Group Working Papers I, 1*, 71-82.

Bradlow, A. R., Torretta, G. M., & Pisoni, D. B. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication, 20*(3-4), 255-272.

Crandell, C. C., & Smaldino, J. J. (2002). Room acoustics and auditory rehabilitation technology. In J. Katz, R. F. Burkard & L. Medwetsky (Eds.), *Handbook of clinical audiology* (5th ed., pp. 607-630). Philadelphia, Pa. ; London: Lippincott Williams & Wilkins.

Fairbank, G. (1960). *Voice and articulation drillbook* (2nd ed.). New York: Harper & Row.

Ferguson, S. H., & Kewley-Port, D. (2007). Talker differences in clear and conversational speech: Acoustic characteristics of vowels. *Journal of Speech, Language, and Hearing Research, 50*(5), 1241-1255.

Fischer-Jorgensen, E. (1967). Phonetic analysis of breathy (murmured) vowels in Gujarati. *Indian Linguistics, 28*, 71-139.

French, N. R., & Steinberg, J. C. (1947). Factors governing the intelligibility of speech sounds. *The Journal of the Acoustical Society of America, 19*(1), 90-119.

Fukazawa, T., el-Assuooty, A., & Honjo, I. (1988). A new index for evaluation of the turbulent noise in pathological voice. *Journal of the Acoustical Society of America, 83*(3), 1189-1193.

Gerratt, B. R., & Kreiman, J. (2004). Voice quality, perceptual evaluation of. In R. D. Kent (Ed.), *The MIT encyclopedia of communication disorders* (pp. 78-80). Cambridge, Massachusetts: The MIT Press.

Gordon-Salant, S. (1986). Recognition of natural and time/intensity altered CVs by young and elderly subjects with normal hearing. *Journal of the Acoustical Society of America, 80*(6), 1599-1607.

Gordon-Salant, S. (1987). Effects of acoustic modification on consonant recognition by elderly hearing-impaired subjects. *Journal of the Acoustical Society of America, 81*(4), 1199-1202.

Hanson, W., & Emanuel, F. W. (1979). Spectral noise and vocal roughness relationships in adults with laryngeal pathology. *Journal of Communication Disorders, 12*(2), 113-124.

Harrell, R. W. (2002). Puretone evaluation. In J. Katz, R. F. Burkard & L. Medwetsky (Eds.), *Handbook of clinical audiology* (5th ed., pp. 71-87). Philadelphia, Pa. ; London: Lippincott Williams & Wilkins.

Heman-Ackah, Y. D., Heuer, R. J., Michael, D. D., Ostrowski, R., Horman, M., Baroody, M. M., et al. (2003). Cepstral peak prominence: A more reliable measure of dysphonia. *Annals of Otology, Rhinology and Laryngology, 112*(4), 324-333.

Henton, C. G., & Bladon, R. A. W. (1985). Breathiness in normal female speech: Inefficiency versus desirability. *Language and Communication, 5*(3), 221-227.

Hillenbrand, J. (1987). A methodological study of perturbation and additive noise in synthetically generated voice signals. *Journal of Speech and Hearing Research, 30*(4), 448-461.

Hillenbrand, J. (1988). Perception of aperiodicities in synthetically generated voices. *Journal of the Acoustical Society of America, 83*(6), 2361-2371.

Hillenbrand, J., Cleveland, R. A., & Erickson, R. L. (1994). Acoustic correlates of breathy vocal quality. *Journal of Speech and Hearing Research, 37*(4), 769-778.

House, A. S., Williams, C. E., Heker, M. H., & Kryter, K. D. (1965). Articulation-testing methods: Consonantal differentiation with a closed-response set. *The Journal of the Acoustical Society of America, 37*, 158-166.

Huffman, M. K. (1987). Measures of phonation type in Hmong. *Journal of the Acoustical Society of America, 81*(2), 495-504.

Kennedy, E., Levitt, H., Neuman, A. C., & Weiss, M. (1998). Consonant-vowel intensity ratios for maximizing consonant recognition by hearing-impaired listeners. *Journal of the Acoustical Society of America, 103*(2), 1098-1114.

Kent, R. D. (1992). Introduction. In R. D. Kent (Ed.), *Intelligibility in speech disorders : theory, measurement, and management* (Vol. 1, pp. 1-10). Amsterdam ; Philadelphia: J. Benjamins Pub.

Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America, 87*(2), 820-857.

Koenig, L. L. (2000). Laryngeal Factors in Voiceless Consonant Production in Men, Women, and 5-Year-Olds. *Journal of Speech, Language, and Hearing Research, 43*(1-5), 1211-1228.

Krause, J. C., & Braida, L. D. (2002). Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility. *Journal of the Acoustical Society of America, 112*(5), 2165-2172.

Krause, J. C., & Braida, L. D. (2004). Acoustic properties of naturally produced clear speech at normal speaking rates. *Journal of the Acoustical Society of America, 115*(1), 362-378.

Kreiman, J., Gerratt, B. R., & Berke, G. S. (1994). The multidimensional nature of pathologic vocal quality. *Journal of the Acoustical Society of America, 96*(3), 1291-1302.

Kreiman, J., Gerratt, B. R., Kempster, G. B., Erman, A., & Berke, G. S. (1993). Perceptual evaluation of voice quality : review, tutorial, and a framework for future research. *Journal of Speech and Hearing Research, 36*(1), 21-40.

Ladefoged, P. (1983). The linguistic use of different phonation types. In D. M. Bless & A. J. H. (Eds.), *Vocal fold physiology: Contemporary research and clinical issues* (pp. 351-360). San Diego: College-Hill Press.

Liu, H. M., Tsao, F. M., & Kuhl, P. K. (2005). The effect of reduced vowel working space on speech intelligibility in Mandarin-speaking young adults with cerebral palsy. *Journal of the Acoustical Society of America, 117*(6), 3879-3889.

Lundy, D. S., Roy, S., Casiano, R. R., Xue, J. W., & Evans, J. (2000). Acoustic analysis of the singing and speaking voice in singing students. *Journal of Voice, 14*(4), 490-493.

Maryn, Y., Roy, N., De Bodt, M., Van Cauwenberge, P., & Corthals, P. (2009). Acoustic measurement of overall voice quality: A meta-analysis. *Journal of the Acoustical Society of America, 126*(5), 2619-2634.

Mendes, A. P., Rothman, H. B., Sapienza, C., & Brown, W. S. (2003). Effects of vocal training on the acoustic parameters of the singing voice. *Journal of Voice, 17*(4), 529-543.

Michaelis, D., Fröhlich, M., & Strube, H. W. (1998). Selection and combination of acoustic features for description of pathologic voices. *Journal of the Acoustical Society of America, 103*(3), 1628-1639.

Milenkovic, P. (2001). TF32. Madison, Wisconsin.

Monsen, R. B. (1978). Toward measuring how well hearing-impaired children speak. *Journal of Speech and Hearing Research, 21*(2), 197-219.

Omori, K., Kacker, A., Carroll, L. M., Riley, W. D., & Blaugrund, S. M. (1996). Singing power ratio: Quantitative evaluation of singing voice quality. *Journal of Voice, 10*(3 2), 228-235.

Osguthorpe, J. D., & Nielsen, D. R. (2006). Otitis externa: Review and clinical update. *American Family Physician, 74*(9), 1510-1516.

Picheney, M. A., Durlach, N. I., & Braida, L. D. (1985). Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech. *Journal of Speech and Hearing Research, 28*, 96-103.

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1985). Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech. *Journal of Speech and Hearing Research, 28*, 96-103.

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1986). Speaking clearly for the hard of hearing. II: Acoustic characteristics of clear and conversational speech. *Journal of Speech and Hearing Research, 29*(4), 434-446.

Radish Kumar, B., Bhat, J. S., & Prasad, N. (2009). Cepstral analysis of voice in persons with vocal nodules. [Article in Press]. *Journal of Voice*, 1-3.

Reetz, H., & Jongman, A. (2009). *Phonetics: Transcription, production, acoustics, and perception*. Chichester, U.K. ; Malden, MA: Blackwell.

Robb, M., & Chen, Y. (2008). A note on vowel space in Mandarin accented English. *Asia Pacific Journal of Speech, Language, and Hearing, 11*(3), 175-188.

Sapienza, C. M., & Ruddy, B. H. (2009). *Voice disorders*. San Diego: Plural Publishing.

Shoji, K., Regenbogen, E., Jong Daw, Y., & Blaugrund, S. M. (1992). High-frequency power ratio of breathy voice. *Laryngoscope, 102*(3), 267-271.

Smith, L. Z., & Levitt, H. (1999). Consonant enhancement effects on speech recognition of hearing-impaired children. *Journal of the American Academy of Audiology, 10*(8), 411-421.

Södersten, M., Hertegård, S., & Hammarberg, B. (1995). Glottal closure, transglottal airflow, and voice quality in healthy middle-aged women. *Journal of Voice, 9*(2), 182-197.

Stouten, V., & Van hamme, H. (2009). Automatic voice onset time estimation from reassignment spectra. *Speech Communication, 51*(12), 1194-1205.

Sundberg, J. (1987). *The science of the singing voice*. DeKalb, Ill.: Northern Illinois University Press.

Theodore, R. M., Miller, J. L., & DeSteno, D. (2009). Individual talker differences in voice-onset-time: Contextual influences. *Journal of the Acoustical Society of America, 125*(6), 3974-3982.

Titze, I. R. (1994). *Principles of voice production*. Englewood Cliffs, N.J.: Prentice Hall.

Titze, I. R. (2000). *Principles of voice production*. Iowa City: National Center for Voice and Speech.

Torre III, P., & Barlow, J. A. (2009). Age-related changes in acoustic characteristics of adult speech. *Journal of Communication Disorders, 42*(5), 324-333.

Urbaniak, G. C., Plous, S., & Lestik, M. (2010). Research Randomizer (Version 4.0): Social Psychology Network.

Venema, T. (2006). *Compression for clinicians* (2nd ed.). Clifton Park, NY: Thomson Delmar.

Volaitis, L. E., & Miller, J. L. (1992). Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories. *Journal of the Acoustical Society of America, 92*(2 I), 723-735.

Watts, C., Barnes-Burroughs, K., Estis, J., & Blanton, D. (2006). The singing power ratio as an objective measure of singing voice quality in untrained talented and non-talented singers. *Journal of Voice, 20*(1), 82-88.

Wolfe, V. I., Martin, D. P., & Palmer, C. I. (2000). Perception of Dysphonic Voice Quality by Naive Listeners. *Journal of Speech, Language, and Hearing Research, 43*(3), 697-705.

Wu, K., & Childers, D. G. (1991). Gender recognition from speech. Part I: Coarse analysis. *The Journal of the Acoustical Society of America, 90*(4), 1828-1840.

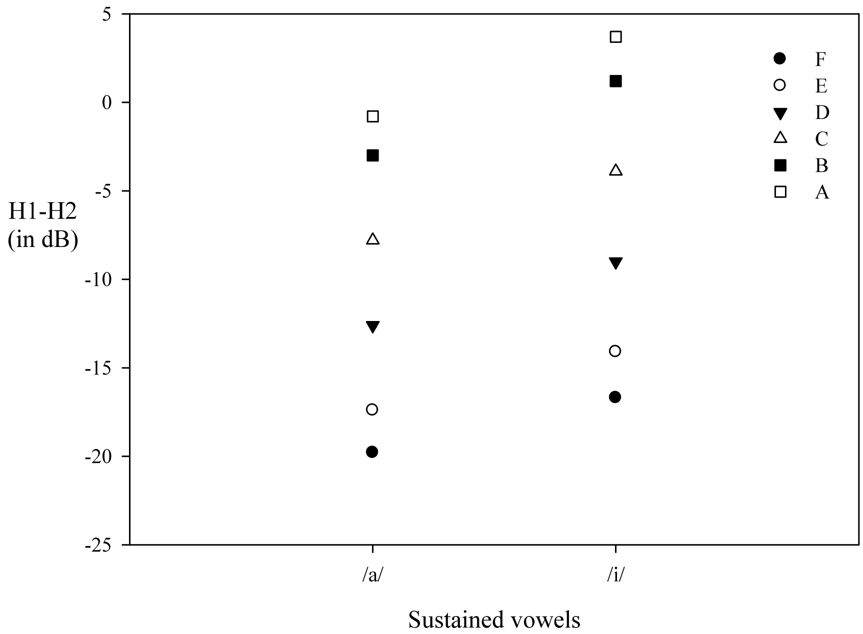
Appendix 1

The first six sentences from the "Rainbow Passage" with the words selected for measurement of acoustic parameters in bold face.

1. When the sunlight strikes raindrops in the air, they act as a prism and form a rainbow.
2. The rainbow is a division of white light into many beautiful **colours**.
3. These **take** the shape of a **long** round **arch**, with its path high above, and its **two** ends apparently beyond the horizon.
4. There is, according to legend, a boiling **pot** of gold at one end.
5. **People** look, but no one ever finds it.
6. When a man looks for something beyond his **reach**, his friends say he is looking for the pot of gold at the end of the rainbow.

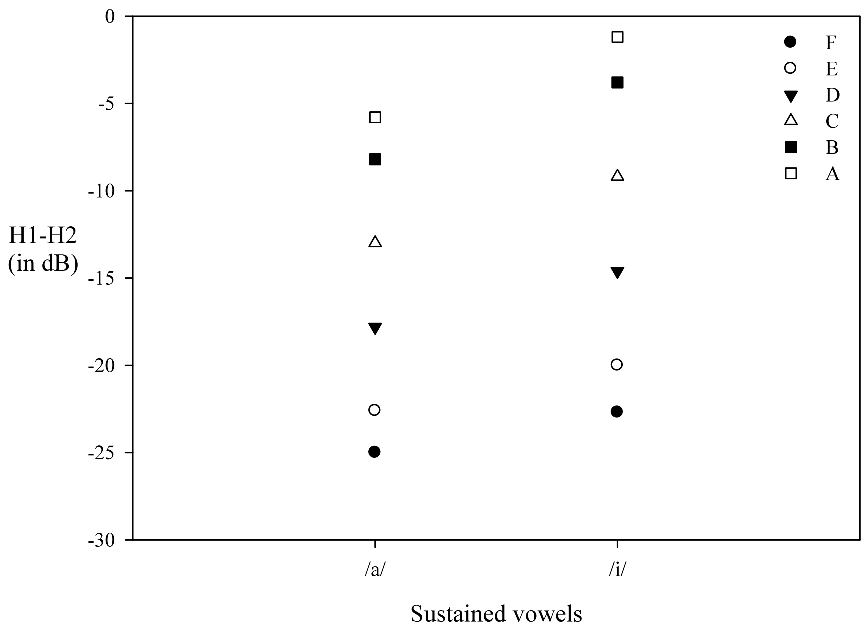
Appendix 2

Ranges for H1-H2 levels of perceptual study "male sustained vowels". H1-H2 amplitude difference values of stimuli fell within the range of A to B (Level 1), B to C (Level 2), C to D (Level 3), D to E (Level 4), and E to F (Level 5). Levels 1 to 5 represent increases in H1-H2 amplitude difference with Level 1 being the lowest level.



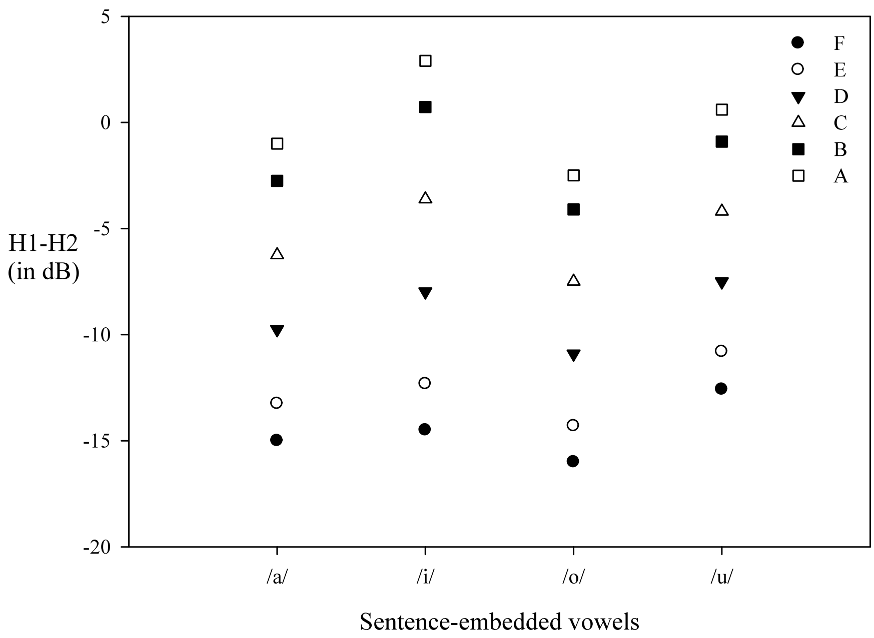
Appendix 3

Ranges for H1-H2 levels of perceptual study “female sustained vowels”. H1-H2 amplitude difference values of stimuli fell within the range of A to B (Level 1), B to C (Level 2), C to D (Level 3), D to E (Level 4), and E to F (Level 5). Levels 1 to 5 represent increases in H1-H2 amplitude difference with Level 1 being the lowest level.



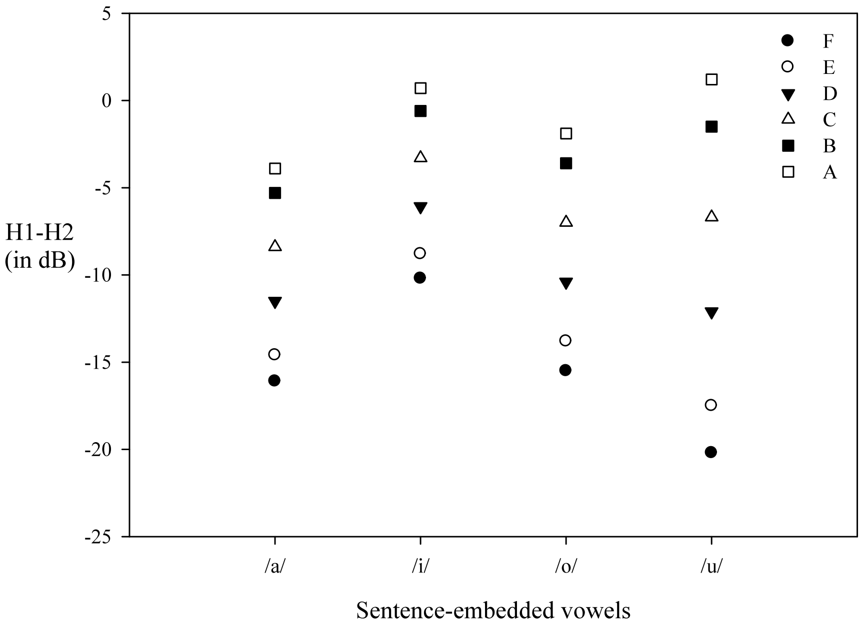
Appendix 4

Ranges for H1-H2 levels of perceptual study "male sentence-embedded vowels". H1-H2 amplitude difference values of stimuli fell within the range of A to B (Level 1), B to C (Level 2), C to D (Level 3), D to E (Level 4), and E to F (Level 5). Levels 1 to 5 represent increases in H1-H2 amplitude difference with Level 1 being the lowest level.



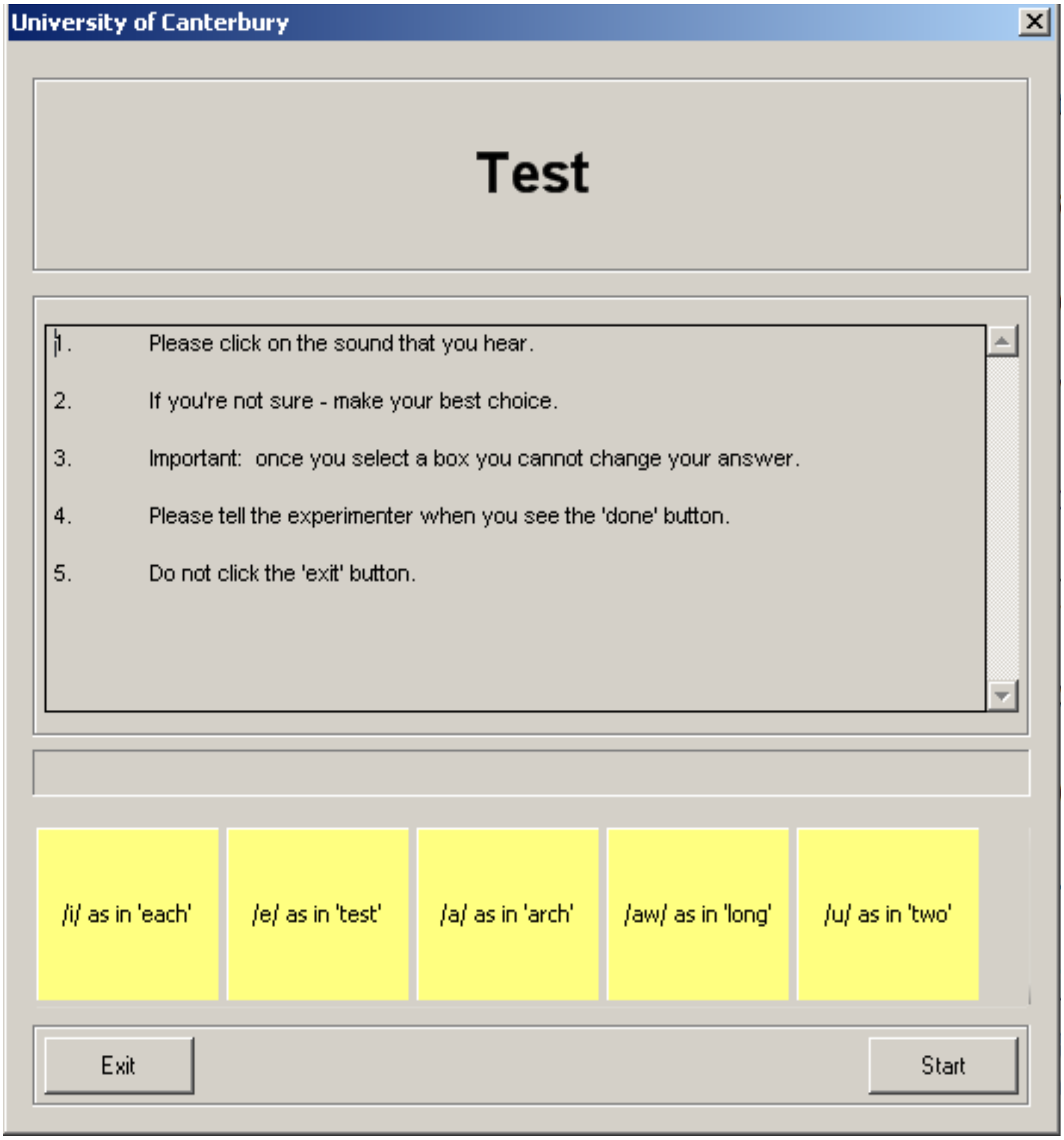
Appendix 5

Ranges for H1-H2 levels of perceptual study “female sentence-embedded vowels”. H1-H2 amplitude difference values of stimuli fell within the range of A to B (Level 1), B to C (Level 2), C to D (Level 3), D to E (Level 4), and E to F (Level 5). Levels 1 to 5 represent increases in H1-H2 amplitude difference with Level 1 being the lowest level.



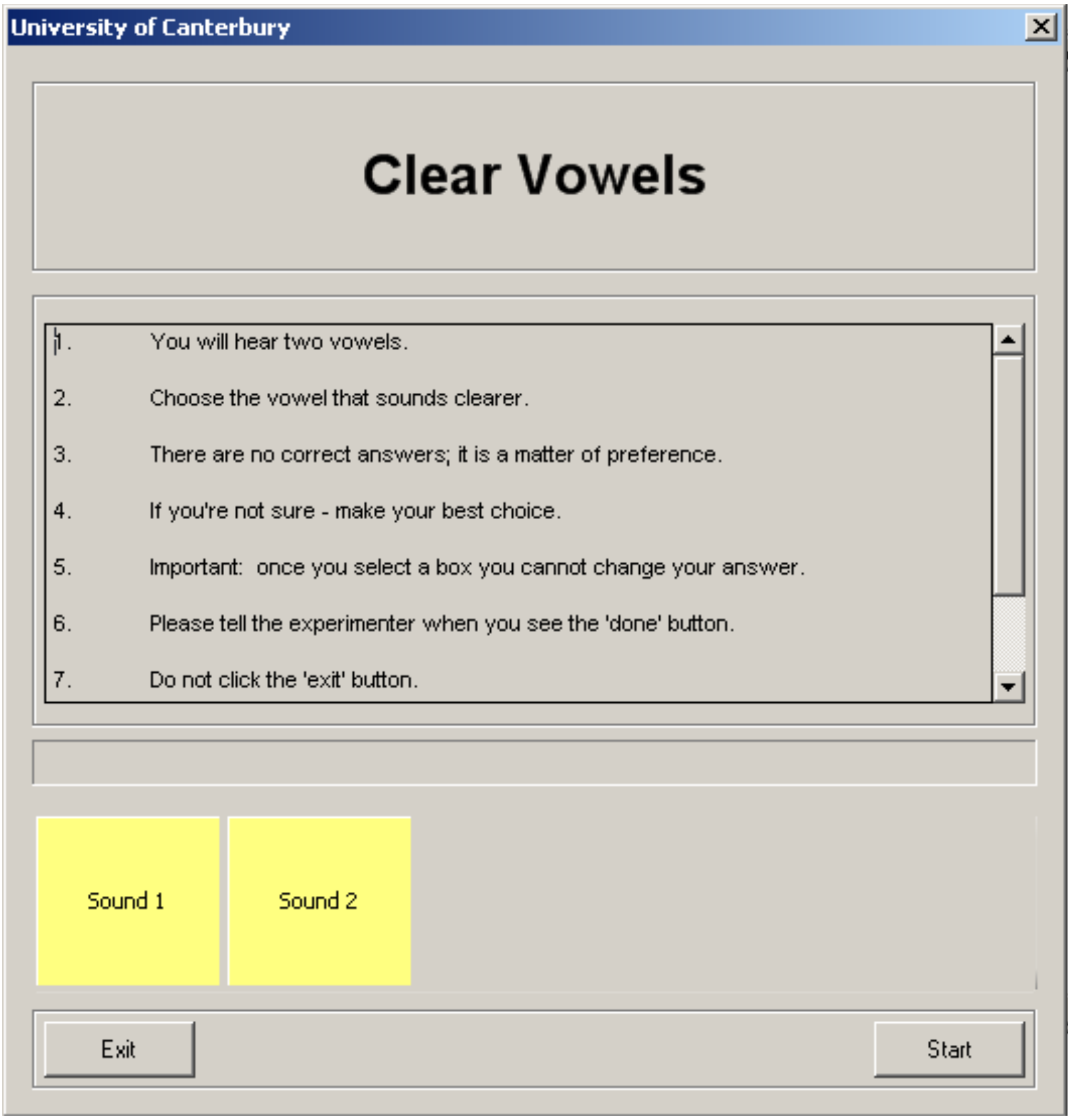
Appendix 6

Programme interface screen for the "vowel identification" task.



Appendix 7

Programme interface screen for the "perceived as clearer" task.



Appendix 8

Vowel Space

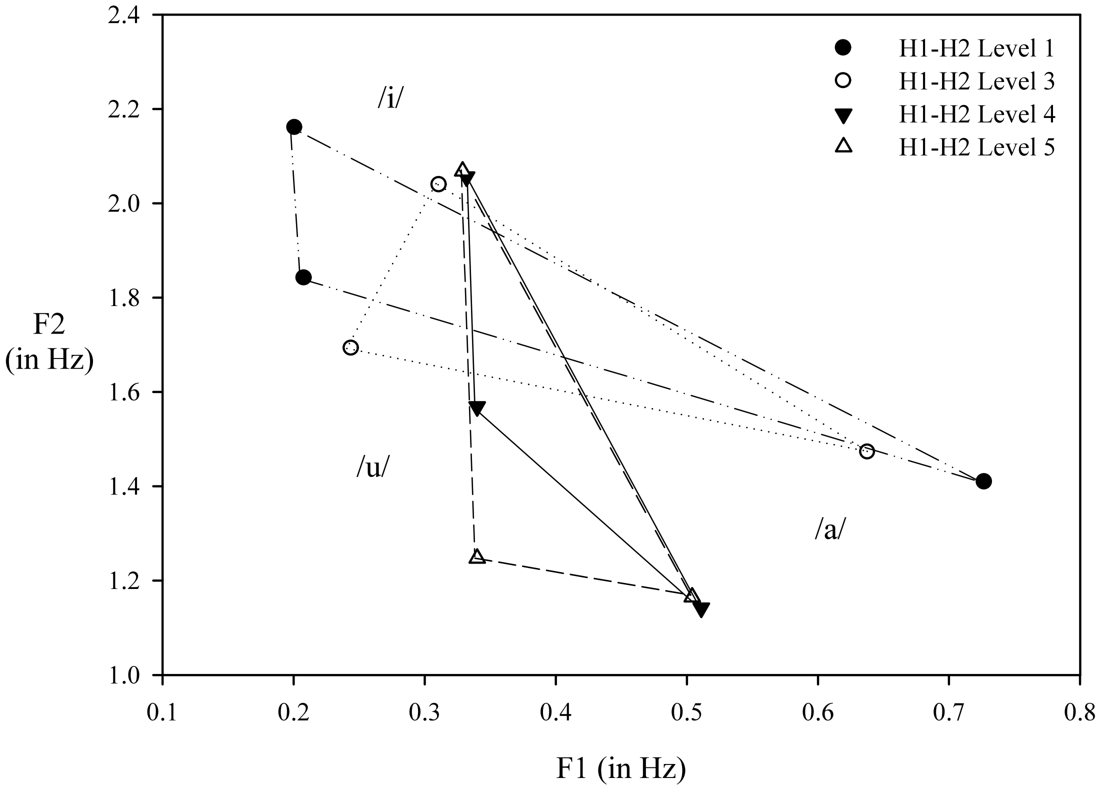
F1-F2 acoustic vowel space plots were constructed from F1 and F2 measures measured from male and female “sentence-embedded vowels” separately. The stimuli were selected on the basis of levels of H1-H2 amplitude difference and gender so that each triangle was made up of three vowels of the same level and there was one triangle for each level and a set of triangles for each gender. The vowel stimuli for each triangle was not necessarily produced by any one individual talker but may have resulted from contributions from two or more voices. Contrary to expectation of increasing acoustic vowel space (associated with improved intelligibility) with increasing levels of H1-H2 amplitude difference, the vowel spaces for male produced stimuli went in order of increasing size: Level 4, Level 5, Level 3, Level 1 (see Appendices 9 and 11). For the female produced stimuli the acoustic vowel space areas were similarly contrary going in order of increasing size: Level 5, Level 1, Level 4, Level 3 (see Appendices 10 and 11).

Further F1-F2 acoustic vowel space plots were constructed from F1 and F2 values measured from vowel stimuli selected so that the vowels for each triangle did come from an individual talker. The level of the H1-H2 amplitude difference was used to select the individual talkers for these vowels sot that there was one triangle for each level and one set for each gender. The stimuli for vowels /i/ and /u/ for each triangle had H1-H2 amplitude difference levels as close as the level for the vowel /a/ as was possible given the constraint of having been produced by the same talker. For male talkers the acoustic vowel space area, contrary to expectations, went in order of increasing size: Level 5, Level 3, Level 1, Level 4 (see Appendices 12 and 14). For female talkers the acoustic vowel space area followed a pattern of, in order of increasing size: Level 5, Level 1, Level 3, Level 4 (see Appendices 13 and 14).

Overall the vowel space plots showed that vowel space changes as a function of the level of H1-H2 amplitude differences proved difficult to interpret.

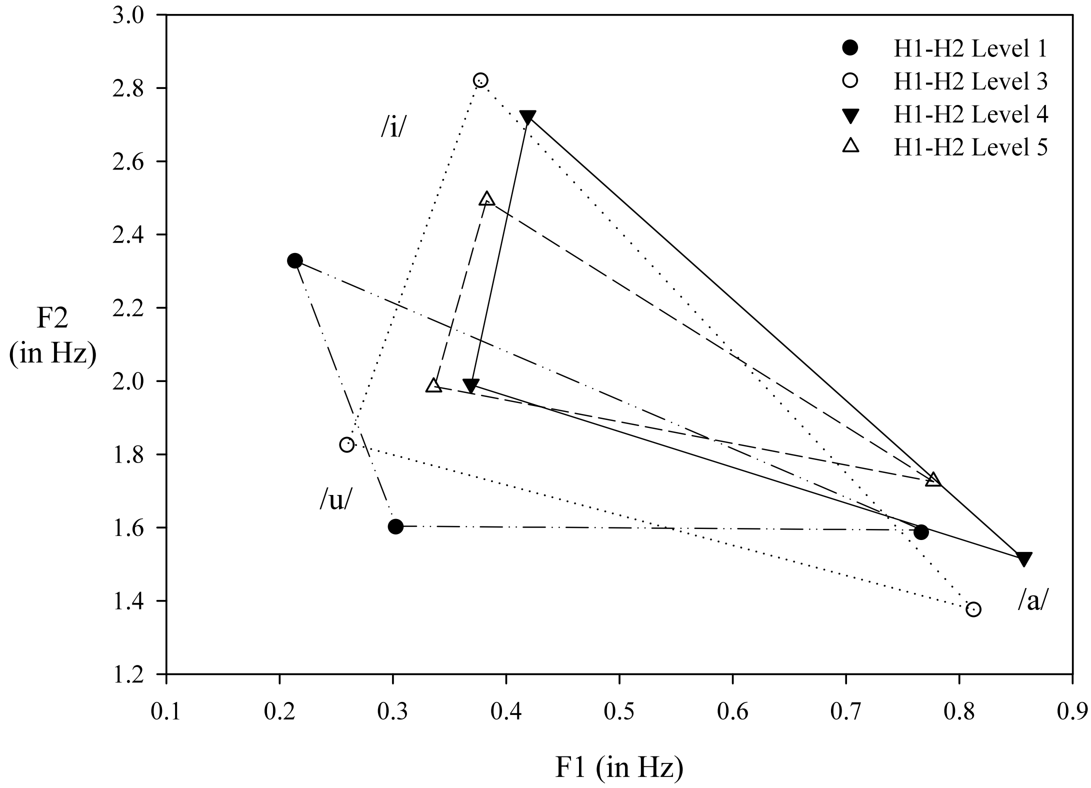
Appendix 9

Vowel space derived from pooled F1 and F2 values from "male sentence-embedded vowels". Vowels were selected on the basis of the H1-H2 amplitude difference. Each vowel space triangle is potentially made up of contributions from different talkers. Levels L1 to L5 represent increases in H1-H2 amplitude difference with L1 being the lowest level.



Appendix 10

Vowel space derived from pooled F1 and F2 values from "female sentence-embedded vowels". Vowels were selected on the basis of the H1-H2 amplitude difference. Each vowel space triangle is potentially made up of contributions from different talkers. Levels L1 to L5 represent increases in H1-H2 amplitude difference with L1 being the lowest level. All of the vowels in these vowel space triangles were used in the perceptual study.



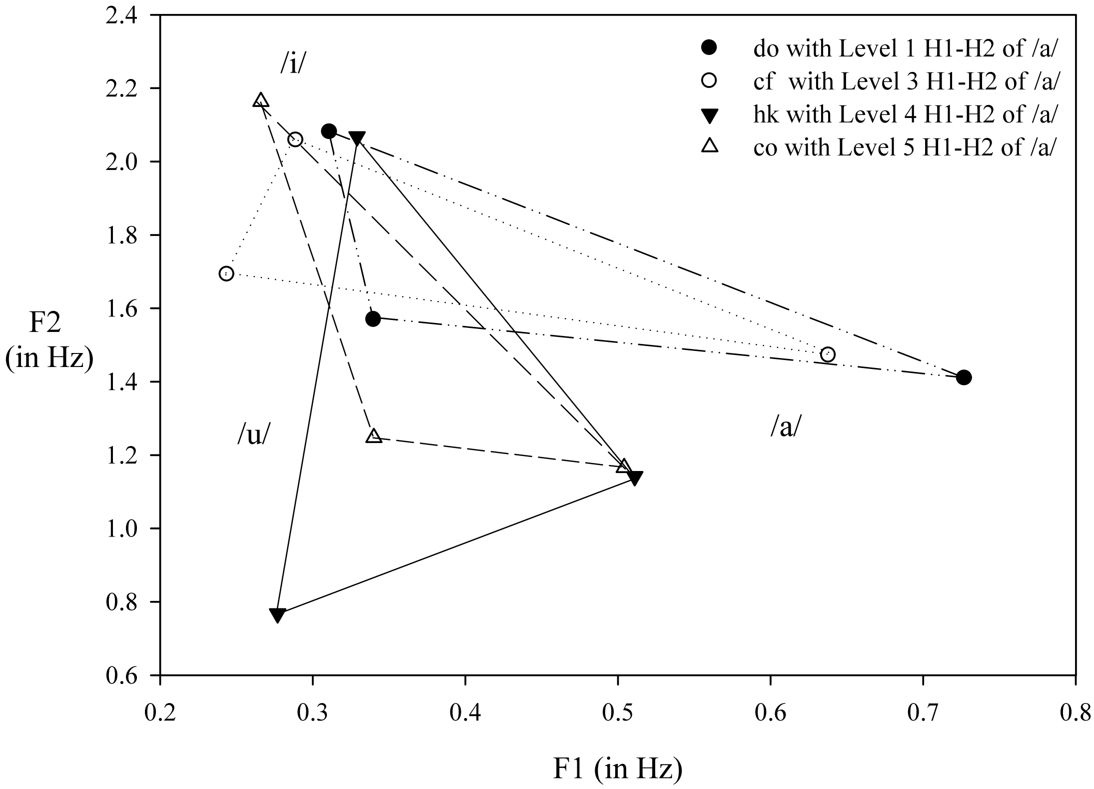
Appendix 11

Acoustic vowel space areas. Acoustic vowel space areas measured from F1-F2 acoustic vowel space plots constructed from pooled F1 and F2 values from "male and female sentence-embedded vowels". The different stimuli for each triangle were determined by the H1-H2 amplitude difference levels. Each vowel space triangle is potentially made up of contributions from different talkers.

Level of H1-H2	Gender	Vowel space area (kHz^2)
1	Male	0.081
	Female	0.168
3	Male	0.076
	Female	0.302
4	Male	0.040
	Female	0.191
5	Male	0.067
	Female	0.119

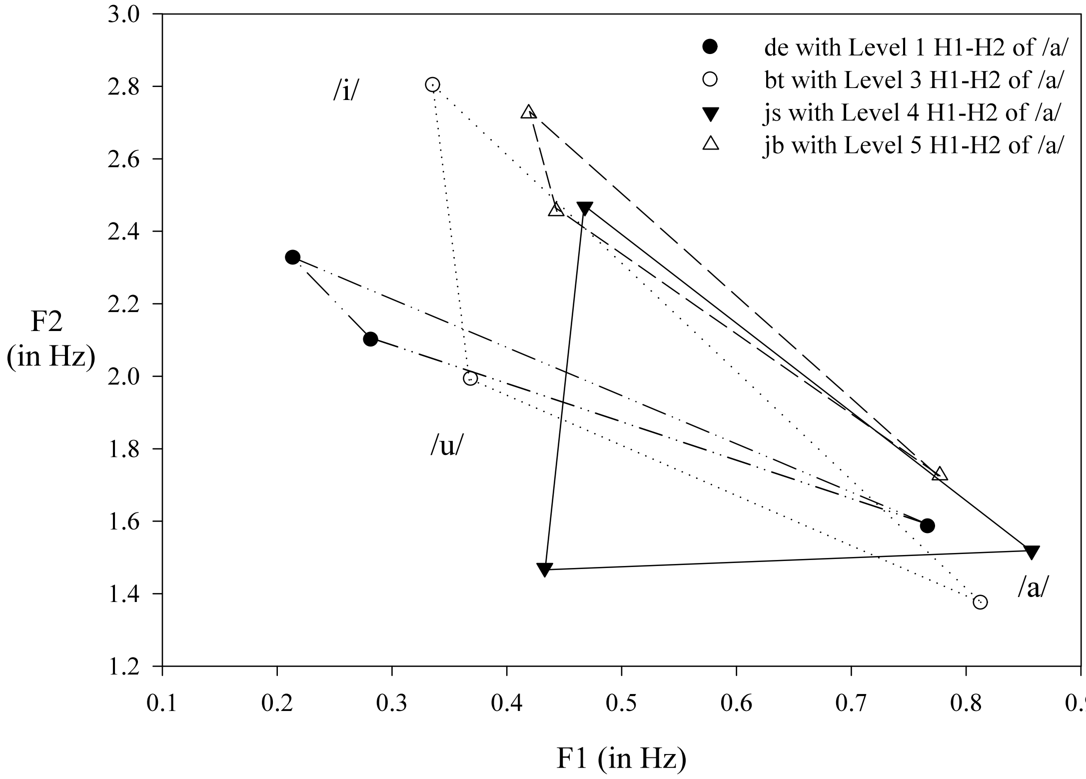
Appendix 12

Vowel space derived from individual talker's F1 and F2 values from "male sentence-embedded vowels". Individual talkers were selected on the basis of the H1-H2 amplitude difference level of the vowel /a/. The H1-H2 levels of the vowels /i/ and /u/ did not necessarily conform to those of the vowel /a/. Levels L1 to L5 represent increases in H1-H2 amplitude difference with L1 being the lowest level. All of the vowels in these vowel space triangles were used in the perceptual study.



Appendix 13

Vowel space derived from individual talker's F1 and F2 values from "female sentence-embedded vowels". Individual talkers were selected on the basis of the H1-H2 amplitude difference level of the vowel /a/. The H1-H2 levels of the vowels /i/ and /u/ did not necessarily conform to those of the vowel /a/. Levels L1 to L5 represent increases in H1-H2 amplitude difference with L1 being the lowest level. All of the vowels in these vowel space triangles were used in the perceptual study.



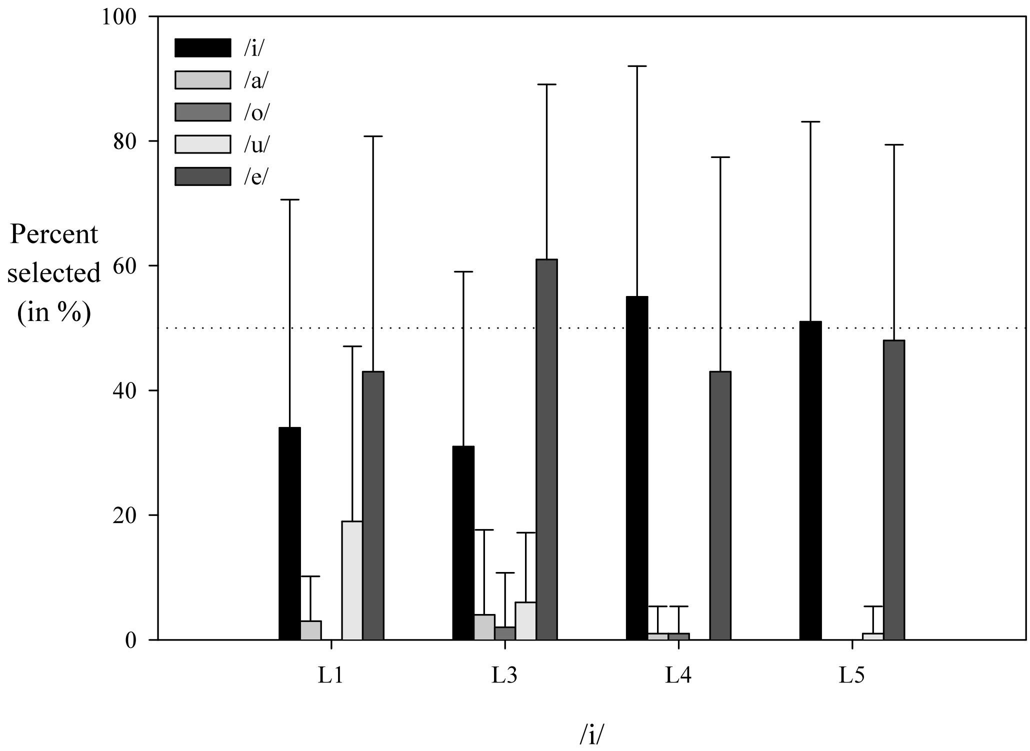
Appendix 14

Acoustic vowel space areas. Acoustic vowel space areas measured from F1-F2 acoustic vowel space plots constructed for selected male and female talkers. The different levels were determined by the H1-H2 amplitude difference of the vowel /a/. The levels of H1-H2 amplitude difference measured from the other vowels /i, u/ making up the triangle for each talker were not necessarily the same level as for the vowel /a/ for each talker.

Level of H1-H2 for /a/	Gender	Glottal closure	Vowel space area (kHz^2)
1	Male	Incomplete	0.097
	Female	Incomplete	0.037
3	Male	Complete	0.077
	Female	Complete	0.170
4	Male	Complete	0.142
	Female	Complete	0.211
5	Male	Complete	0.072
	Female	Complete	0.036

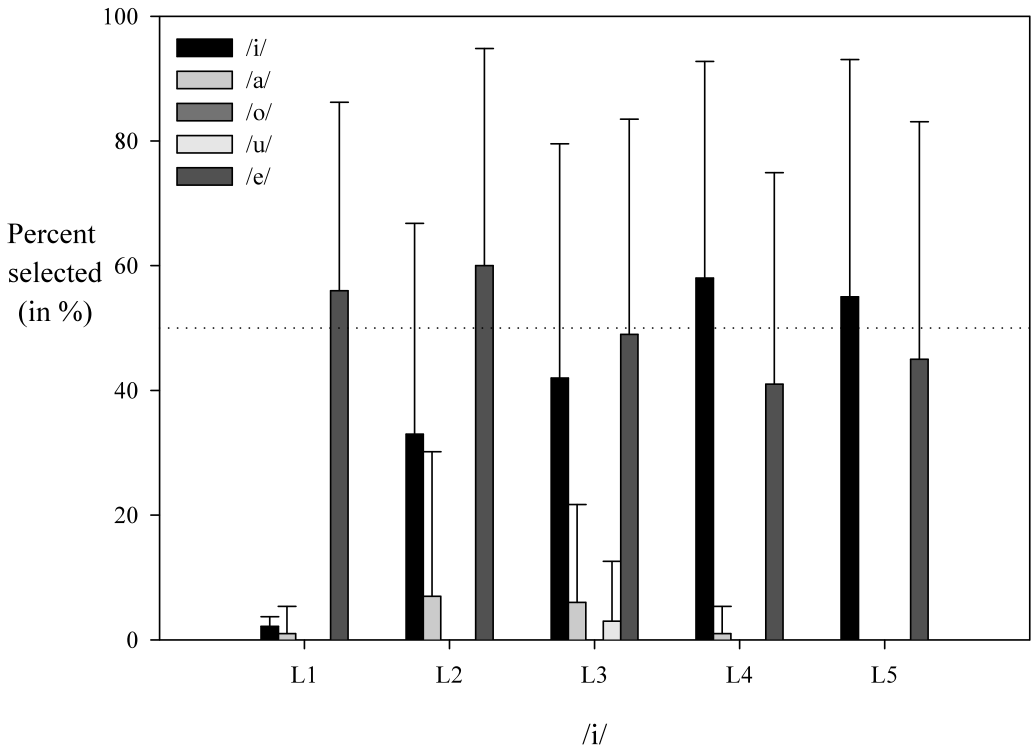
Appendix 15

Averaged listener responses, correct and incorrect, to the male vowel /i/. Means and standard deviations from averaged listener responses, to the vowel /i/, be they correct /i/ responses or incorrect /a, o, u, e/ responses (n = 100), for the first set of “male sentence-embedded vowels” used in the “vowel identification” task. Levels L1 to L5 represent increases in H1-H2 amplitude difference with L1 being the lowest level.



Appendix 16

Averaged listener responses, correct and incorrect, to the female vowel /i/. Means and standard deviations from averaged listener responses, to the vowel /i/, be they correct /i/ responses or incorrect /a, o, u, e/ responses (n = 100), for the first set of "female sentence-embedded vowels" used in the "vowel identification" task. Levels L1 to L5 represent increases in H1-H2 amplitude difference with L1 being the lowest level.



Appendix 17

Averaged listener responses, correct and incorrect, to the male and female vowel /a/ for Level 2 of H1-H2 amplitude difference. Means and standard deviations from averaged listener responses, to the vowel /a/, be they correct /a/ responses or incorrect /i, o, u, e/ responses (n = 100), for the first set of "male and female sentence-embedded vowels" used in the "vowel identification" task.